

# The heat is on

At December's climate-change meeting, everyone can agree on one thing: it is make-or-break time.

**N**ext month's United Nations Climate Change Conference in Bali, Indonesia, is charged with drawing up a clear and convincing road map that will lead to a robust international climate-change agreement to succeed the Kyoto Protocol. That is a momentous challenge but, given the right approach from participants, not an insurmountable one.

Evidence is fast mounting that time is running out for nations to unite in a credible response to climate change. The International Energy Agency said last week that energy-related emissions of carbon dioxide are set to grow from 27 gigatonnes in 2005 to 42 gigatonnes by 2030 — a rise of 56%. Other estimates project even higher growth, and also reveal, alarmingly, that 'carbon intensity' — the level of carbon emissions required to sustain a given level of economic activity — is actually growing again.

Bad faith still clouds the implementation of the 1997 Kyoto Protocol, which expires in 2012. US withdrawal from the agreement and the treaty's subsequent failure to make much noticeable impact on its signatories' emissions leave ample scope for ugly and fruitless recrimination.

Yet the timing of the Bali meeting is, politically, rather more auspicious than was the case at Kyoto ten years ago. A new politics of climate change is unfolding in Australia (see page 336) and the United States (see page 340), two of the nations where opposition to effective international action has been strongest (see also pages 333 and 342). This is the context for a new international agreement that must transcend the Kyoto accord by setting in motion a vision that is more broadly based and more ambitious than the cap-and-trade principle of the original agreement.

## If the cap fits

The road map that will emerge from Bali need not cling to cap-and-trade as its mainstay. Indeed, cap-and-trade has yet to be fully tested as a mechanism for reducing emissions. It was originally conceived to deal with sulphur emissions from power stations — an issue in which the costs and benefits are relatively small — but doubts are emerging about its ability to cope with the far larger issue of greenhouse-gas emissions. Nevertheless, it should not be abandoned, but other approaches, including the direct imposition of carbon taxes, may ultimately prove to be equally as potent in raising the cost of burning fossil fuels.

In addition to striking the right balance between Kyoto-type targets and fresh approaches to global action on climate change, the Bali road map will have to cover territory that the previous accord was unable to reach. For example, it will have to set terms for the direct involvement of developing nations in the international push to reduce carbon emissions. That means the participation of nations such as India, Brazil and China — the last of which has, the International Energy Agency believes, this year surpassed the United States as the world's largest emitter of carbon dioxide.

The poorest countries cannot be expected to be part of this push here, but they must benefit from assistance for their efforts to adapt to

the effects of climate change. Several wealthy nations, such as the Netherlands, are forging ahead with sophisticated adaptation strategies. But for the most vulnerable societies, adaptation to climate change ultimately boils down to poverty alleviation. Such a requirement must co-exist with the politically awkward fact that the new accord must take into account substantial contributory factors that were excluded from Kyoto — including emissions from air transport and the huge impact of deforestation.

A long-term international commitment to reduced emissions will also involve far greater collaboration between nations on the research and development programmes needed to come up with more alternative-energy sources. Already, market forces are pulling a diverse range of solar and wind options through to more cost-effective operation. But there are major energy problems, such as carbon capture and sequestration, that will require substantial public-sector investment to move forward.

**"The road map to emerge from Bali will have to cover territory that the Kyoto Protocol was unable to reach."**

## Thinking big

All of this is a tall order for the representatives of more than 180 nations expected to gather at Bali on 3–14 December. A great deal will hinge on the approach taken by the two largest emitters, China and the United States.

There will be a temptation for China to assume a stance as defender of the poor nations against the malign interests of the rich. It would certainly be politically feasible for it to point to the failure of the United States to fulfil its earlier agreement that developed nations should act first, and to argue that there is no moral case for China to do so when its per capita emissions are so much lower than those of other countries.

Fortunately, there are signs that China knows better than to adopt this dead-end strategy. There are clear signs that Beijing is fully aware of the threat posed to China's future development by climate change and, as usual, its leadership will be taking the long view. Despite the rampant construction of coal plants in the past few years, China knows that its energy balance has to change, and it can use the Bali meeting to assert itself as a global power, offering to lead, rather than to follow.

The outlook for the US delegation is less promising. Its negotiators will include ideologues who will stop at nothing to derail the humble progress the rest of the world has managed to make through the Kyoto accord. The only thing that can save them from themselves is President Bush's possible desire to end his reign on a slightly more positive note than the salutary one on which he began it, with his March 2001 decision that the United States would withdraw from the Kyoto agreement. The next American president is, thankfully, likely to take a more constructive approach to global warming after 2009. The Bali meeting is the global community's chance to prepare for the opening that this should provide. ■

# Enhancing, not cheating

A broad debate about the use of drugs that improve cognition for both the healthy and the ill is needed.

**O**f all the arguments levelled against taking drugs for human enhancement, the idea that it is cheating has least power. Yes, when organized competition or formal testing of abilities is the name of the game, drug-based strengthening is questionable and regulations against it need to be adhered to.

More debatable are arguments by opponents of drug-based enhancement that it is cheating against oneself. "Personal achievements impersonally achieved are not truly the achievements of persons," said the report *Beyond Therapy* by the US President's Council on Bioethics, chaired at the time by ethicist Leon Kass. Yet imagine if a researcher could improve his or her ability to memorize the postulated connections in a complex signalling pathway central to tumour development, or if a musician could improve his or her concentration and deliver a better performance on the night. Far from cheating on themselves or others, they would be delivering a higher return on their investment of effort, and indeed on society's investment in them. We all benefit.

What is sure is that opponents of enhancement are, to a degree, whistling in the wind. They raise other spectres — unfairness of access (although today's enhancing dose is cheaper than a cup of coffee), possibilities of employer coercion and the loss of human dignity or of the 'natural' — but ultimately, to little avail. Many healthy people still opt for chemical enhancements of all sorts, as suppliers of cosmetics and some pharmaceuticals know well. Such actions betoken an ethical argument on the other side: the pursuit of personal liberty.

Studies on healthy volunteers have shown the cognitive effects of enhancing drugs to be mild, but sufficient to be considered helpful. The pills with least risky side effects seem to be methylphenidate (prescribed for attention-deficit disorder) and modafinil (prescribed for narcolepsy). Studies, mostly in the United States, have documented usage of drugs for cognitive enhancement by 5–15% of students, and anecdotes abound of use by postdocs and academics.

**"Studies have documented usage of drugs for cognitive enhancement by 5–15% of students."**

As described in a valuable discussion paper from the British Medical Association ([www.bma.org.uk/ap.nsf/Content/CognitiveEnhancement2007](http://www.bma.org.uk/ap.nsf/Content/CognitiveEnhancement2007)), these drugs have an honourable track record in helping the afflicted — and they could do yet more to that end. The risks of long-term use are not well understood (although that also applies to many approved drug therapies). And the ethical issues are not to be ignored, although the document lacks a description of who stands to make money from enhancement drugs. Appropriately, it reviews the issues and sets out topics for debate, rather than advocating use or restriction of enhancers.

The debate itself is not new, but it has been confined to relatively small circles. Yet the use of cognitive-enhancing drugs by both the ill and the healthy — and those in the substantial grey zone overlapping both — is already more widespread than is generally appreciated. The lack of debate on regulation by nations or by institutions, such as universities, is increasingly problematic. And those in favour of chemically induced cognitive enhancement for the healthy need to develop their scientific and ethical case, because it will be all too easy for the ethically conservative to hold sway over political leaders, most of whom would prefer to wish this particular challenge away. ■

# Prescription for change

Health research in Italy is in desperate need of a fresh start.

**T**hese are painful times for Livia Turco, the Italian health minister. A member of the centre-left Democratic Party, Turco has been caught in a web of power politics that has led her to nominate Enrico Garaci to serve a third term as president of the ISS, an important, publicly funded health-research institute in Rome. The nomination is seen by many as problematic because Garaci has not fully embraced the open and competitive peer review that Italian research policy must adopt if it is to compete more effectively with other scientific powers of comparable size.

On 6 November, the Italian Senate's health committee took the unusual move of rejecting the nomination. Legally, Turco is obliged to take note of the Senate's view — and also that of the chamber of deputies, a committee of which approved the nomination on 24 October — but she does not have to follow it. By withdrawing the nomination, she may lose political face, but by insisting on it, she will undermine her government's main objective, which is to cajole Italian governance into a new era of meritocracy and openness.

The ISS is in some ways Italy's equivalent of the Pasteur Institute in Paris. It employs around 1,500 scientists who work in areas such as vaccines, stem cells and genomics, and its €100 million (US\$145 million) annual budget is mostly absorbed by salaries — although the institute also coordinates some extramural projects.

The way in which Garaci has administered these projects has often upset other senior scientists. Their discontent is currently focused on €3 million allocated to stem-cell research this year. Stem-cell researchers have complained to Turco in a letter to which she has not replied. Newspapers have pointed out that Garaci was a member of the health-ministry committee that helped decide that the ISS would distribute the stem-cell funds. Moreover critics fear that Garaci's own doctrinaire brand of Catholicism — he is a member of the conservative Science and Life group — may prevent the small programme from supporting work that would be permitted under the law, but of which he may personally disapprove.

On balance, Garaci lacks the confidence among his peers that a director of the ISS needs. Turco should withdraw his nomination and follow the procedure adopted successfully by her colleague, research minister Fabio Mussi, in filling top positions. She should set up an independent search committee to draw up a shortlist of candidates from which she can select a nominee, who would then have the full confidence of Italy's biomedical research community. ■

# RESEARCH HIGHLIGHTS

## PARTICLE PHYSICS

### Neutrinos on ice

*Phys. Rev. Lett.* **99**, 171101 (2007)

A large-scale experimental simulation has bolstered an ambitious project to detect neutrinos from space as they hit huge swathes of polar ice.

Neutrinos can provide insights into fundamental physics and stellar nuclear fusion. But because they barely interact with matter, they are very hard to detect.

The project, called ANITA (pictured right), aims to use balloon-borne detectors to scour the Antarctic ice sheets for radio-frequency Cerenkov emissions produced by charged particles kicked up by the incoming neutrinos. But these emissions had not previously been detected in ice.

Using the Stanford Linear Collider to fire neutrinos into a 7.5-tonne block of ice, Peter Gorham of the University of Hawaii and his colleagues have verified that these Cerenkov emissions are indeed produced in ice and can be detected with the balloon instruments.



J. KOWALSKI

## BIOLOGY

### Changed in a flash

*Proc. R. Soc. B* doi:10.1098/rspb.2007.0966 (2007)

A predatory coral-reef fish called the bluestriped fangblenny (*Plagiotremus rhinorhynchus*) swiftly changes colour — from various earth tones to black with an electric-blue stripe (pictured below) — to camouflage itself as a harmless fish species.

The ability to change colour was already known, but Karen Cheney and her colleagues at the University of Queensland in Australia have now shown that it occurs when the fangblenny swims among the fish it mimics — juvenile cleaner fish (*Labroides dimidiatus*). In so doing, the fangblenny can approach fish that are expecting to be cleaned, and then take a bite out of them.

The researchers caught fangblennies from the Great Barrier Reef, and observed them in tanks with both cleaner fish and other types of fish. Their findings are the first to show

that a vertebrate will switch its colour back and forth depending on the presence of the mimicked species.

## ANIMAL BEHAVIOUR

### Breaking the rule

*Proc. R. Soc. B* doi:10.1098/rspb.2007.1056 (2007)

'The island rule' is an old saw of ecology. Small mammals on islands evolve bigger body size than their mainland counterparts; large mammals shrink. The rule recently grabbed attention when people delighted in visions of Flores Island in the Pleistocene, during which miniature humans (*Homo floresiensis*) might have hunted giant rats and lizards.

In the most rigorous test yet of this hypothesis, Shai Meiri of Imperial College, London, and his colleagues looked at 1,184 island-mainland population pairs both separately and in an analysis that took into account their phylogenetic relationships. They found that the island rule does not hold true

across all mammals. It does seem to hold for some groups: mice are often bigger on islands and deer are often smaller. The fact that these groups have many members may have swamped the data set and given rise to an overgeneralization.

## MICROBIOLOGY

### Crowd control

*PLoS Biol.* **5**, e302 (2007).

The shape of the bacterium *Escherichia coli* might have

evolved to cope with life in tight spaces, researchers suggest.

Bacteria can fill tiny cracks and crevices with ordered colonies united by an active cell-to-cell communication system. Andre Levchenko of Johns Hopkins University in Baltimore, Maryland, and Alex Groisman of the University of California, San Diego, and their colleagues observed *E. coli* cells as they colonized chambers in a microfluidic device, which is designed to force fluids into tiny spaces.

Simulations based on their observations showed that the length-to-width ratio of the bacteria enables cells to avoid a traffic jam or being damaged by other cells while leaving the confining chambers of the device. The organization of the colony also helps to maintain nutrient flow to every cell.

## GENOMICS

### A head-scratcher

*Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.0706756104 (2007)

Fungi associated with human dandruff have surprising similarity to other skin-colonizing organisms, genome sequences reveal. This insight might lead to targets for treatment of various skin disorders.

Tom Dawson of Proctor and Gamble and his collaborators sequenced the genomes of *Malassezia globosa* and, less thoroughly, *Malassezia restricta*. They then compared the sequences with those of a closely related plant pathogen, *Ustilago maydis*, and a distantly related human pathogen, *Candida albicans*, which causes thrush.



K. CHENEY



Despite an overall genomic similarity, *U. maydis* and *Malassezia* differ in the proteins they secrete. *Malassezia* secretes proteins that metabolize lipids and fats from the host's skin that are similar to those of the distantly related *C. albicans*, probably because they evolved to live in similar niches.

Researchers also discovered that *Malassezia* might be able to reproduce sexually.

## NANOTECHNOLOGY

### Tiny tunes

*Nano Lett.* 10.1021/nl0721113 (2007); *Nano Lett.* 10.1021/nl0714839 (2007)

Two groups of researchers have constructed radios out of nanotubes.

Alex Zettl and his colleagues at the University of California at Berkeley exposed a nanotube with a charged tip to radio waves. The waves cause the tip to vibrate, and when the frequency of the incoming wave matches the resonance frequency of the tube, the device can convert the transmission into an audible signal.

Meanwhile, Chris Rutherglen and Peter Burke of the University of California, Irvine, developed their own nanotube radio by exploiting the nonlinear current-voltage characteristics of a single-walled carbon nanotube that was fixed to electrodes at both ends.

Zettl's team picked up the Beach Boy's song Good Vibrations on their machine, and Burke's team broadcast harp music from Turlough O'Carolan, a blind musician from the seventeenth century.

## IMMUNOLOGY

### It's not you, it's your protein

*Nature Immunol.* doi:10.1038/ni1527 (2007)

The success of bone-marrow transplants could depend on a mouse protein called Sirpa. Jayne Danska at the Hospital for Sick Children Research Institute in Toronto, Canada, and her colleagues explain why one strain of mice with a mutation that severely compromises its immune system can accept bone marrow, whereas other strains with the same mutation kill the invading cells.

The team's special strain of mice boasts a version of Sirpa that silences the immune system's onslaught. Adding this version to immune cells in a Petri dish allows bone-marrow cells grown with them to flourish and produce blood cells. Found on host macrophages, Sirpa recognizes a protein on bone-marrow cells called CD47.

The researchers hope to determine whether the 10 or more versions of Sirpa found in

human populations can explain transplant failures between seemingly matched patients.

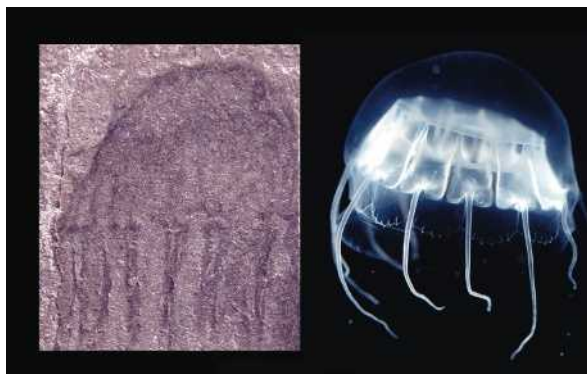
## PALAEONTOLOGY

### Ancient jellyfish

*PLoS ONE* 2, e1121 (2007)

The discovery of jellyfish fossils dating from about 500 million years ago (pictured below left) shows that the complex invertebrates evolved much earlier than was previously thought. Four types of jellyfish, peeled from the shale of northwestern Utah, are the earliest known definitive examples of these animals. Previously, the earliest jellyfish fossils came from Pennsylvania shale quarries and were dated to about 320 million years ago.

Lead author Pauly Cartwright of the University of Kansas in Lawrence says that the new fossils from the Marjum Formation reinforce the view that relatives of today's drifting marine life (pictured below right) also lived during the Cambrian period.



B. LIEBERMAN; K. RASKOFF

## NANOCHEMISTRY

### Gene boxes

*J. Am. Chem. Soc.* 129, 13376–13377 (2007)

DNA is emerging as an ideal material for construction of nano-objects because it can be programmed to self-assemble into double-stranded segments according to its base sequence. But building three-dimensionally with DNA has generally been cumbersome and messy. Making objects that are topologically complex, such as cubes, has required many synthetic steps and has given only a low yield.

Faisal Aldaye and Hanadi Sleiman of McGill University in Montreal, Canada, have now developed a versatile way to make various DNA polyhedral nanostructures at a stroke. They create prefabricated polygonal rings of single-stranded DNA, which can then be efficiently linked into cage-like shapes with the addition of complementary linking strands. One of these cages can even change shape.

## JOURNAL CLUB

Clive R. Bagshaw

University of Leicester, UK

**A biochemist is excited by a universal glue for molecular biology.**

Investigating the dynamic properties of proteins at the level of a single molecule allows insight into properties that are masked in ensemble studies. I have often found that the hardest part of such studies is immobilizing the molecule on a silica surface in a 'permanent' way that retains the molecule's function.

The proteins we investigate are usually prepared with a His-tag — comprising 6+ engineered histidine residues — that binds, via a chelated nickel ion, to nitriloacetic acid (NTA), aiding purification on an NTA affinity column. Immobilization through this tag would therefore be an attractive option. But alas, this is only partly successful using the standard NTA group because proteins have a significant probability of detaching from the silica support on the timescale of minutes.

But surely a chemist somewhere has improved on this technology? Thanks to Google, I found the work of Jacob Piehler who, in 2005, introduced tris-NTA, a cyclam ring with three groups attached to it. Tris-NTA shows a thousand-fold higher affinity for His-tags in the presence of nickel than NTA and a dissociation half-life of many hours.

Piehler and colleagues have gone on to exploit this technology as a general means of attaching fluorophores to His-tagged proteins and, most recently, as a convenient way of specifically conjugating proteins to streptavidin (A. Reichel et al. *Anal. Chem.* doi:10.1021/ac0714922; 2007).

The streptavidin-biotin complex is another widely-used 'glue' in biotechnology, but the use of an intermediate tris-NTA-biotin adaptor broadens its application to His-tagged proteins and renders the attachment reversible on addition of excess imidazole. I look forward to using this technology in our single-molecule studies, for which such a reversible glue has the same appeal as a Post-It note.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>



## NEWS

# Pakistan's universities take up protest

Scientists and students in Pakistan have joined countrywide protests against the state of emergency, the suspension of the country's constitution and the sacking of its judiciary announced by the president, General Pervez Musharraf, on 3 November.

Thousands of staff and students are continuing to demonstrate at some of the country's top universities, among them the University of the Punjab, Lahore University of Management Sciences and the GIK Institute of Engineering Sciences and Technology in Topi.

There are also daily protests at the country's flagship Quaid-i-Azam University in Islamabad. On Tuesday, police tried to arrest some 500 protesting students and faculty at the university. But, according to news reports, vice-chancellor Qasim Jan stepped in and insisted that the protest continued.

This is the first time the scientific community has opposed military rule in such large numbers, says Pervez Hoodbhoy, chair of physics at Quaid-i-Azam University. "This is an unexpectedly big number, because for 20 years, since student unions were banned by General Muhammad Zia-ul-Haq [the last military ruler], Pakistani universities have been quite indifferent to national politics."

Since taking power in a coup in 1999, General Musharraf has assumed the twin roles of head of state and head of the army. But he has also carved a reputation as an economic and social



Mass demonstrations by university staff and students now occur daily in Pakistan.

liberal, and a bulwark against the al-Qaeda terrorist network. He has allowed independent media organizations to publish and broadcast. And Internet access has expanded from 40 cities in 1999 to 2,000 in 2006, with one in two Pakistanis having a mobile phone.

Musharraf also presided over a massive university building programme, and introduced a

tenure-track system for university professors. The numbers of PhDs emerging from universities has more than doubled to 700 annually, with plans for a further doubling in the next four years.

But like all of Pakistan's previous military dictators (Musharraf is the fourth), he is having trouble letting go of the reins of power. With

F. NAEEM/AFP/GETTY

# Merck settles Vioxx lawsuits for \$4.85 billion

Three years after it pulled its blockbuster painkiller Vioxx from the market, Merck has agreed to pay \$4.85 billion to settle nearly 27,000 lawsuits that claim the arthritis drug caused heart attacks and strokes.

The settlement, although specifying that Merck "does not admit causation or fault", is in effect the legal expression of an admission by the company of a medical problem when it withdrew the \$2.5-billion-a-year drug in September 2004 after five years on the market. It was prompted to do so when patients taking Vioxx in a Merck-sponsored clinical trial showed a doubled risk of heart attack and stroke (R. Bresalier *et al.* *N. Engl. J. Med.* 352, 1092–1102; 2005). At the time, the company promised to fight in court every lawsuit brought against it.

**"Merck does not admit causation or fault."**

The settlement agreed on 9 November is a qualified positive for Merck. The New Jersey company's Vioxx liability was initially pegged by Wall Street at tens of billions of dollars. What's more, the deal will go ahead if only 85% of eligible claimants agree to it, lowering the risk that thousands of cases will continue to drag on in court.

Both sides claim victory. Plaintiffs' attorney Mark Lanier of the Lanier Law Firm in Houston, Texas, applauded the deal as "a huge opportunity for those who have been hurt by the drug to receive a fair settlement" and said he would counsel the 1,006 plaintiffs he represents to accept it.

Bruce Kuhlik, Merck's general counsel, called the settlement "a product of the success that we've had in the litigation so far".

Despite a stunning \$253-million judgment against Merck in the first Vioxx trial (see *Nature* 436, 1070; 2005) — later reduced to around \$26 million under Texas damage law — Merck has won 11 Vioxx cases so far, whereas plaintiffs have prevailed in only 5.

Those numbers reflect the difficulty of proving causation in individual cases, regardless of the soundness of aggregate findings from large randomized clinical trials, when a drug's adverse effects — and the risk factors that can lead to them — are broadly and independently prevalent in the population. That helped defence lawyers persuade juries that Vioxx wasn't at fault in the heart attacks suffered by plaintiffs such as Elaine Doherty, an overweight 68-year-old with diabetes and high blood pressure.

"What was so clever about the defence

**HAVE YOUR SAY**

Comment on any of our news stories, online.

[www.nature.com/news](http://www.nature.com/news)

elections set for January, some analysts are predicting that he may agree to a power-sharing alliance with opposition leader Benazir Bhutto in which he remains head of state but gives up his role as head of the army, with Bhutto assuming the role of prime minister (for the third time).

Such a scenario could have uncertain consequences for science and education. Bhutto's previous stint as prime minister (1993–96) was characterized by low investment in research and higher education. She was also much criticized by the scientific community for failing to

appoint a minister for science.

**"For 20 years, Pakistani universities have been quite indifferent to national politics."**

Leaders of Pakistan's scientific community are meeting this week at the Academy of Sciences for the Developing World in Trieste, Italy, for a seminar originally designed

to showcase the country's progress in science, technology and higher education. Much of the meeting is now likely to be devoted to speculation on what will happen after the election.

Among the speakers will be Atta-ur-Rahman, chair of Pakistan's Higher Education Commission and a professor of chemistry at the University of Karachi. He told *Nature* that an incoming government had a responsibility not to throw away the past: "I hope the new government will continue on the present path. We have made huge advances in the past several years and it would be a great shame if this were lost."

Ehsan Masood

strategy is that they argued it at the individual level. It certainly drove down the estimates of how much this was going to cost them," says Garret FitzGerald, chair of pharmacology at the University of Pennsylvania in Philadelphia, whose mechanistic studies in humans and mice first pointed to the cardiovascular risks of Vioxx and other drugs in its class.

The Vioxx cases are in stark contrast to the lawsuits brought by people who took Fen-Phen, a popular diet-drug combination withdrawn in 1997. The signature, and relatively uncommon, heart-valve damage caused in those cases left drug-maker Wyeth liable for damages of more than \$21 billion.

Indeed, for scientists and clinicians, the main message of the settlement may be that reliable, personalized lawsuit victories remain as elusive as personalized medicine itself, and for many similar reasons.

Meredith Wadman

## HIV vaccine may raise risk

An HIV vaccine made by Merck that failed in clinical trials may have increased the susceptibility of some trial participants to the AIDS virus, researchers reported last week. The findings have left scientists grappling with the problem of how to handle future trials of vaccines that use similar strategies to stimulate an immune response.

The trial was suspended in September, when preliminary analyses showed that the vaccine failed to protect participants against HIV (see *Nature* 449, 390; 2007). The vaccine consisted of three HIV genes carried by a weakened viral vector. Neither the HIV genes nor the vector, made from a common-cold virus called adenovirus 5, were capable of causing infection.

But it seems that trial participants who had a pre-existing immunity to adenovirus 5 were more likely to become HIV-positive if they received the vaccine. Among male volunteers who had high levels of antibodies against the adenovirus, 21 of 392 became infected in the vaccinated group, but only 9 of 386 were infected in the placebo group. The vaccine did not affect infection rates in men with no adenovirus-5 immunity.

The trial was not designed to examine the relationship between adenovirus-5 immunity and HIV infection, and researchers cannot say whether the result is statistically significant. But they can't ignore the findings, says Anthony Fauci, director of the National Institute of Allergy and Infectious Diseases in Bethesda, Maryland. The start of another HIV-vaccine trial has been pushed back several months to allow more time for scientific review, he says.

For now, researchers are trying to determine whether the vector is to blame or whether those with high adenovirus-5 immunity are at risk for unrelated reasons. Immunologists say it is also theoretically possible that vaccination in those with pre-existing immunity may have triggered an immune response that transiently boosted production of activated CD4 cells — the very immune cells that HIV infects.

That theory leaves several questions unanswered. For example, it is not known how long after vaccination the activated CD4 cells would have remained in the mucosal surfaces where HIV first gains entry, says Fauci.

Researchers are nevertheless

concerned. One possible solution is to limit future trials of adenovirus-5-based vaccines to participants with very low levels of immunity to the vector, says Gary Nabel, director of the Vaccine Research Center at the National Institutes of Health in Bethesda, Maryland. But doing so would close off the trials to a large proportion of the population they aim one day to help: more than 80% of sub-Saharan Africans have significant levels of adenovirus-5 immunity. In addition to the planned HIV-vaccine trial, an Ebola vaccine in development also relies on an adenovirus-5 vector.

But at this stage in vaccine development, researchers are just looking for a proof of concept rather than a vaccine that can be put directly to use, says Lawrence Corey, lead investigator of the HIV Vaccine Trials Network. Alternatives to adenovirus 5 are being developed, and may be substituted once the efficacy of a vaccination strategy has been shown. Nevertheless, vaccine developers will face challenging decisions in light of the new data, he says: "Translating uncertain data into policy is always a difficult issue."

Heidi Ledford



Hopes for a future vaccine against the devastating Ebola virus may have been set back by the results of a clinical trial of an HIV vaccine.

J.-M. BOUJ/AP



## Q&amp;A

# The price of power

The price of oil is hovering at around US\$100 a barrel, a psychologically powerful level that experts and analysts once discussed in purely theoretical terms. **John Deutch**, a chemistry professor at the Massachusetts Institute of Technology in Cambridge and former director of the US Central Intelligence Agency under President Clinton, gives his thoughts on the issue.

## What is the significance of \$100 oil?

The rise in oil prices indicates the inevitable truth that we are using up low-cost energy reserves in the ground. We should expect that over time, on average, oil and gas prices will increase. That increase will cause three things to happen: less oil and gas will be consumed; there will be a shift from oil and gas to renewable energy sources; and, happily, new technologies will become more attractive. I'm quite optimistic.

## What is causing this price rise?

Demand for oil and gas, especially from rapidly growing Asian economies such as India and China, is a real driver. I think that the International Energy Agency projects that oil consumption will go up from 80 or 90 million barrels per day worldwide to about 120 million barrels per day by 2030. The fastest growth will be from Asian economies.

## Is \$100 oil here to stay?

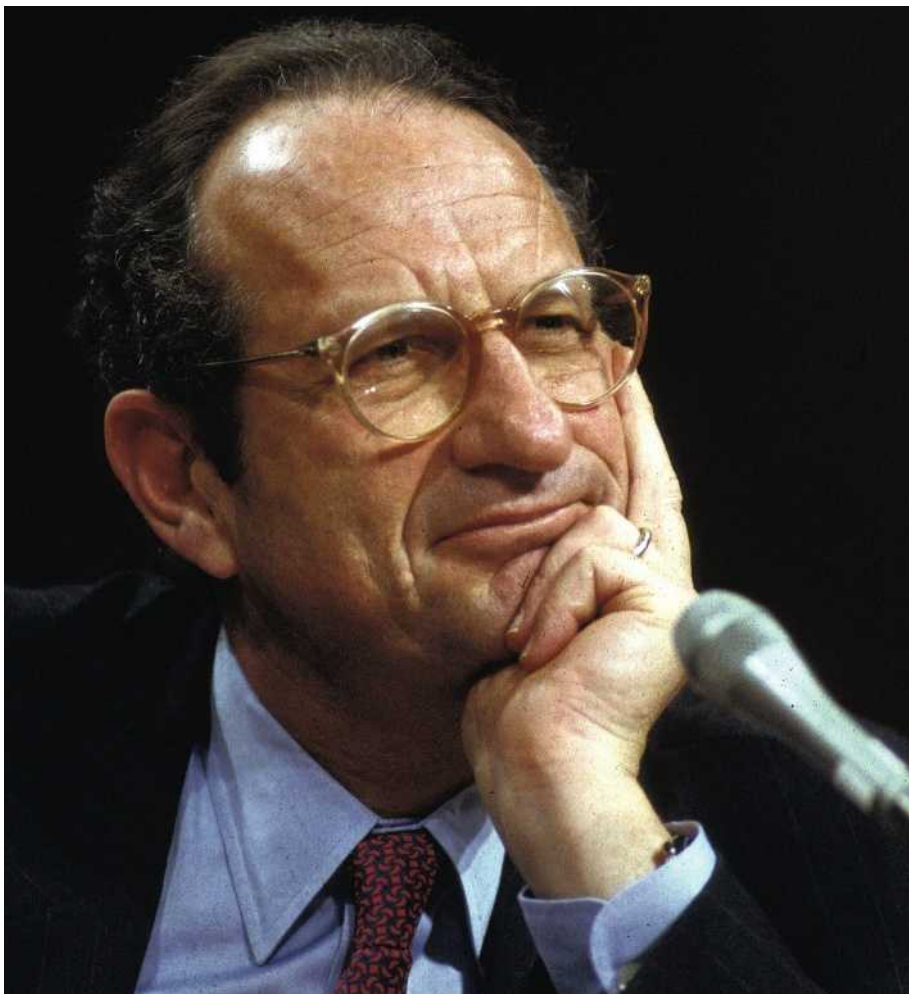
Energy experts know that prices fluctuate. Although I think it is perfectly possible that oil prices will decrease from the present level to as low as \$40 or \$50 per barrel, I don't think we will see it drop to \$20 again. And if you look at it on a decade-by-decade basis, the march of oil prices will be up, and gas prices too.

## Will prices ever get high enough to cause a global recession?

Talented economists point out that oil is less of a major factor in world economy than it was 20 or 30 years ago. But it's certainly the case that as the price goes up, there will be adverse economic consequences for different countries, and some nations will suffer more than most. The small importing countries in Africa and the Caribbean, they will really suffer.

## Are we approaching a peak oil scenario?

Higher (and stable) prices open the economic window for new technologies. There is no [price] threshold; it is gradual. For example, biofuels from biomass are thought to cost \$40–50 per barrel when produced at a commercial scale; synthetic liquids from shale or coal cost perhaps \$50–70 per barrel.



C. KLEPONIS/ZUMA PRESS/NEWS.COM

## How much does US dependence on oil influence international diplomacy?

The energy-security consequences of oil and gas are very heavily dependent on the behaviour of markets and governments. The fact that we depend so much on oil from the Persian Gulf — from Saudi Arabia, Iraq, Iran and Kuwait — leads to questions about these nations' political stability and the role of terrorists. The fact that Iran is supplying 2.9 million barrels of oil per day to world oil markets certainly constrains the freedom of action that Europe and the United States have on other issues of importance, for example on how Iran is pursuing its clandestine nuclear-weapons programme, and on its activities in Iraq.

## How critical is it that a carbon-pricing system be put in place if the economics of energy is to be changed?

I think the issue of a carbon-emissions price, either through a cap-and-trade system or a tax, has a more direct importance and urgency with respect to climate change. Of course it also affects the transition from a petroleum-based market, but those are two really quite separate problems. Environmental constraints must accompany the commercialization of any new technology. For example coal-to-liquid fuels should be produced only if accompanied by carbon capture and sequestration.

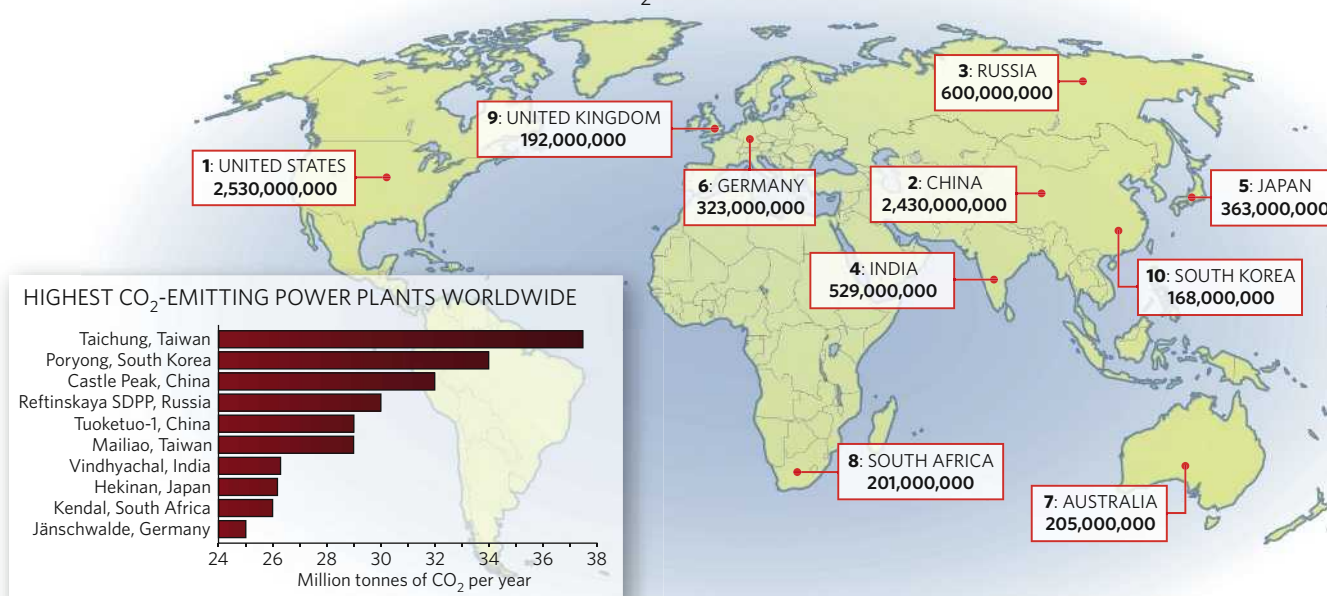
Interview by Jeff Tollefson





**NATURE REPORTS  
CLIMATE CHANGE**  
The news behind the  
science, the science  
behind the news  
[www.nature.com/climate](http://www.nature.com/climate)

## GRAPHIC DETAIL COUNTRIES WITH HIGHEST CO<sub>2</sub>-EMITTING POWER SECTORS (TONNES PER YEAR)



The city of Taichung in Taiwan is home to a power plant that emits more than 37 million tonnes of carbon dioxide into the atmosphere each year, the highest of any plant in the world. Australia produces more carbon dioxide per capita through electricity generation than any other nation. But the US power sector still produces the most carbon dioxide in terms of sheer volume.

These are just a few of the facts and figures available on a new database containing emissions and energy data

from more than 4,000 power companies and 50,000 power plants around the globe. The Carbon Monitoring for Action website was set to be launched this week by the Center for Global Development, an independent think-tank in Washington DC.

With some 8,000 power plants emitting 2.5 billion tonnes of carbon dioxide annually, the United States accounts for a quarter of global emissions, according to the site. China comes a close second at 2.4 billion tonnes,

although its per-capita emissions are less than a quarter of those of the United States. Russia comes a distant third, with 600 million tonnes of carbon dioxide emissions.

The centre describes the database as a public tool that can be used to shed light on the power sector, and hopes it will "speed the shift to less carbon-intensive power generation".

Jeff Tollefson  
[www.carma.org](http://www.carma.org)

# Panel negotiates climate 'synthesis report'

Bathing in the still-warm glow of receiving the 2007 Nobel Peace Prize, the United Nations Intergovernmental Panel on Climate Change is meeting in Valencia, Spain, this week to negotiate the summation of its report on global warming.

The IPCC has issued three scientific reports in the past year, and its fourth 'synthesis report' is scheduled for release on Saturday. The report aims to condense the three volumes — nearly 2,800 pages — into roughly 100 pages for policy-makers and government delegates who will be gathering in Bali, Indonesia, next month for the latest round of global-warming treaty talks (see page 319).

Scientists began arriving in Valencia last week to work out a consensus draft before the arrival of the government delegations.

Stephen Schneider, a climatologist at Stanford University in California, says that the most-recent draft received 2,000 comments from governments. That compares to 5,000 comments from individual scientists on the previous draft.

Given that the report will contain no new science, the challenge will be to get the scientific community and international governments to agree, paragraph by paragraph, on concise



The extent of sea-level rise from polar ice-melt will be hotly debated.

language that lays out the facts without downplaying or overstating the problem at hand.

Many have criticized the process as being too political. The conservation group WWF is already warning that "politically inspired trimming" could unnecessarily dilute the main body of science.

"Some governments will propose emphasizing the more dangerous aspects of the findings. Some will emphasize the uncertainties," says Schneider, who is one of those working on the final draft. "It's our job not to let either one of them go too far."

There will be much after-hours debate over what should be included in the summary, and what can be omitted. The report must also overcome legitimate divides between the three working groups, each of which approached the

science from a different perspective. For instance, the first working group, which based its forecasts on climate models, was necessarily cautious about sea-level rise, saying that sea levels could rise between 18 and 59 centimetres during the twenty-first century. The second working group, however, had to look at the possibility that accelerated warming could cause sea levels to rise much faster than expected.

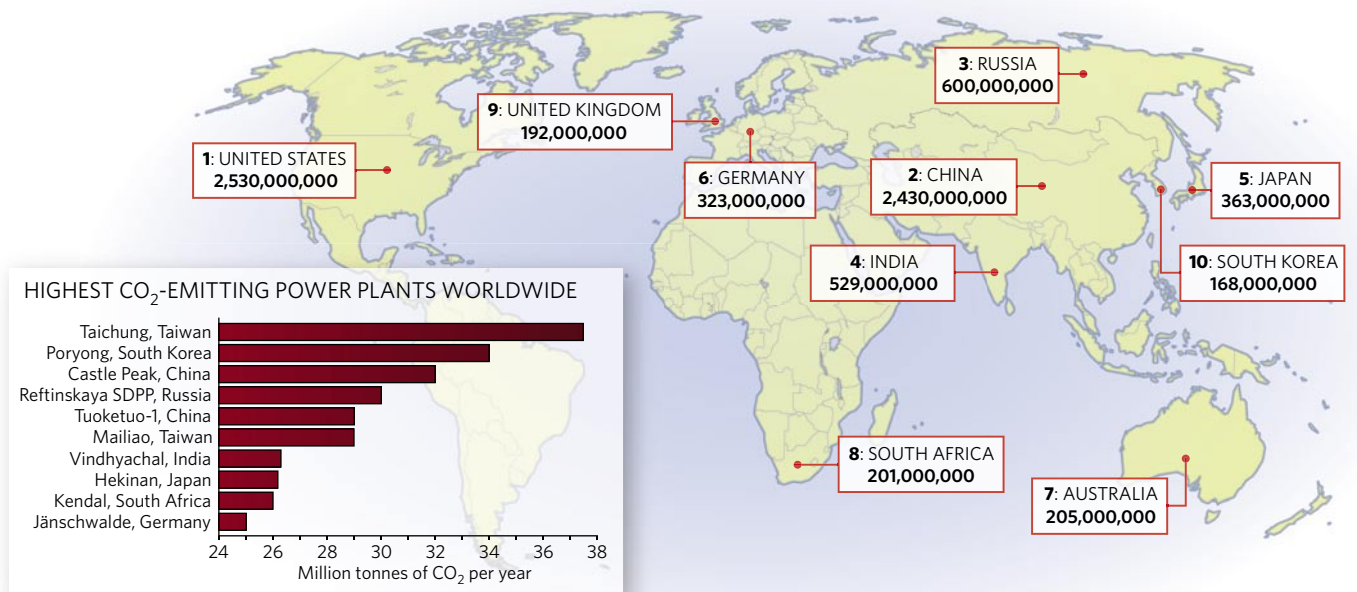
After a year of negotiations, the two groups agreed to a statement that "risks of metres of sea-level rise in centuries cannot be excluded," Schneider says. "That is weaker than working group II wants, stronger than working group I, but it's a synthesis of the reports." The question now, he adds, is whether such language will be accepted by the government delegations.

Jeff Tollefson

J. MCCONNICO/AP

**Correction**

In the Graphic Detail in last week's News (*Nature* **450**, 327; 2007), the label South Korea erroneously pointed to mainland China. The correct version of the map is shown below.

**GRAPHIC DETAIL** COUNTRIES WITH HIGHEST CO<sub>2</sub>-EMITTING POWER SECTORS (TONNES PER YEAR)



**HIGH-ENERGY COSMIC RAYS TRACED TO SOURCE**  
Supermassive black-hole galaxies spin out super-energetic particles.  
[www.nature.com/news](http://www.nature.com/news)

NASA/A. SIMONNET, SONOMA STATE UNIV.



G. PERRONE

## SNAPSHOT Light relief

Luminous jellyfish sculptures fill an exhibition room at the Katonah Museum of Art in New York. The metre-long works by New Orleans

artist Eric Ehlenberger consist of blown-glass bodies attached to three neon-filled tentacles powered by a 12-volt d.c. adaptor plugged into a standard electric outlet. Each jellyfish sells for US\$2,250.

Ehlenberger, who is also an

emergency-room physician, creates abstract and figurative works that focus on marine life, including fish, dolphins and whales. He was recently commissioned by the Seattle Biomedical Research Institute — which is working on a

malaria vaccine — to design a neon sculpture illustrating the feeding posture of anopheline mosquitoes.

The exhibition, called *The Shattering Glass: New Perspectives*, runs until 24 February 2008. **■**  
Nick Thomas

# Brain waves reveal intensity of pain

## SAN DIEGO

Recordings from electrodes in the human brain may offer the first objective way to measure the intensity of pain. Researchers say that they have found a neural signal that correlates with the amount of pain that an individual feels. The signal could be used to refine pain-relief techniques that involve stimulating the brain with electricity, they say.

Single cells have previously been identified in the human brain that are active in pain, but their response is binary, signalling either pain or no pain. Now, Morten Kringelbach of the psychiatry department at the University of Oxford, UK, and his colleagues have identified low-frequency brain waves that emanate from two regions buried deep within the brain when a patient is in pain. The more pain that is experienced, the longer the waves last.

Kringelbach's team recorded activity from

two electrodes positioned in the thalamus and the periaqueductal grey area of 12 awake people who had been undergoing deep-brain stimulation (DBS) for chronic pain. During the recording, the team touched either a painful or pain-free area of the patients' bodies and had patients rate their pain every minute. The duration of the waves — dubbed "pain spindles" — correlated with how intensely the patients felt their pain. "It is an objective measure that correlates with a subjective measure," says Kringelbach, who presented the findings at the Society for Neuroscience meeting in San Diego, California, last week.

This signal could help refine DBS for such patients, he says. Currently, the electrodes fire signals continuously, even when the patient doesn't need pain relief, which wears down the

implanted battery. "We could have a stimulator that picks up this neural signature of pain and only starts sending signals at that point," Kringelbach says.

**"It is an objective measure that correlates with a subjective measure."**

"It would be great to have a 'signature of pain'," says Allan Basbaum, a neuroscientist at the University of California, San Francisco. It is not clear from this preliminary study whether these spindles disappear when painkillers, anaesthetics or electrical stimulation are used, he adds.

The next step is to try to record the signals with a non-invasive technique, such as magnetoencephalography, which measures electrical activity in the brain via changes in the magnetic field. This could allow pain monitoring in a much broader range of patients. **■**

Kerri Smith



# Light wormholes could wire space invisibly

A prescription for wormholes that transmit light invisibly between remote regions has been devised by a team of researchers. These 'electromagnetic wormholes' are the light-based equivalent of space-time wormholes, the staples of science fiction that permit time travel and create short cuts across intergalactic space. They could one day be used in three-dimensional (3D) video-display units wired up with invisible light-carrying cables.

In the electromagnetic wormholes described by Allan Greenleaf of the University of Rochester in New York state and his colleagues, light would disappear in one place and reappear in another, conveyed along channels that cannot be seen from the outside<sup>1</sup>. They remain theoretical at this stage, but the technologies for making them already exist. They draw on the same ideas and methods as those used recently to make invisibility shields<sup>2</sup>. Greenleaf's team envisages a slew of possible applications for their wormholes, in areas ranging from information technology to biomedicine (see 'What to do with a wormhole'). They would, for example, provide perfectly insulated 'optical cables' that totally shield the world outside from the electromagnetic field of the light within.

"It's a cool idea," says physicist Ulf Leonhardt at the University of St Andrews in Scotland,

who specializes in this kind of manipulation of light. The implication, he says, is that "you don't see that two parts of space are connected by the wormhole until you look through it" — rather like looking into your coffee cup and seeing the street outside.

The wormholes would be made from 'metamaterials', structures engineered to interact with light in odd ways. The building blocks of metamaterials — the equivalent of atoms in normal materials — are little electrical circuits, typically made from loops and coils of wire on circuit boards. These interact with the electromagnetic

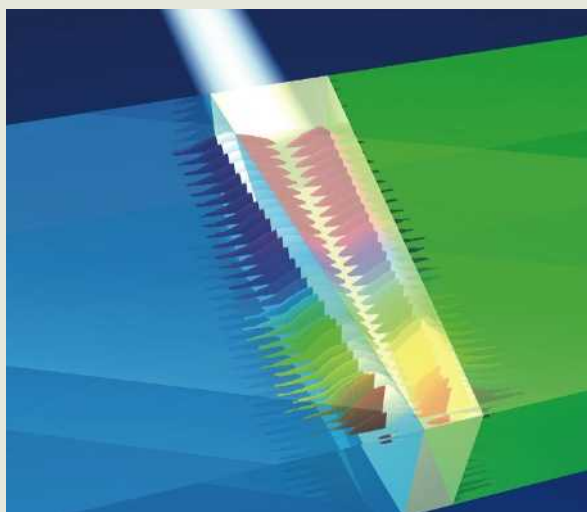
field of a light beam to create unusual optical effects, such as a negative refractive index — the light is bent the 'wrong' way as it passes through.

Last year, Leonhardt and a team led by David Smith of Duke University in Durham,

North Carolina, independently proposed that metamaterials might be used to make an invisibility shield that bends light smoothly around an object placed inside it<sup>3,4</sup>. Smith's group subsequently made such a shield that works at microwave frequencies<sup>2</sup>.

The shielding in that case wasn't perfect, and worked only within a flat plane and for a very narrow frequency range. Moreover, making an invisibility shield for visible light is harder, because the components of the metamaterial

**"You don't see that two parts of space are connected by the wormhole until you look through it."**



## Trapped rainbow

Another use for optical 'metamaterials' is proposed by Tsakmakidis *et al.* on page 397. They show that in theory a metamaterial with a negative refractive index can be used to bring light to a standstill. Such 'slow' — in this case stationary — light has been produced before in materials with strange optical properties, but those have generally

been exotic, for example, ultracold gases of metal atoms, or have had limitations such as producing only modest slowing or working only in narrow wavelength bands.

But the researchers' device works at room temperature, and uses a fabric that could easily be incorporated into light-based (photonic) logic circuits. The light, passing through a

metamaterial waveguide of tapering thickness, is stopped at a position that depends on the wavelength, so that a many-coloured beam gets spread out into a 'trapped rainbow' (left). The effect could be exploited for storing optical information encoded in light pulses, or for introducing time delays into photonic circuits to help synchronize signals. **P. B.**

## WHAT TO DO WITH A WORMHOLE

- **OPTICAL CABLES:** the wormhole could act as a perfect optical fibre that is invisible from the outside — but only at a single wavelength, the one the tube is designed to carry.
- **OPTICAL COMPUTERS:** data-processing elements for conducting logical operations on light signals could be placed inside the wormhole, so that all one would see of the computer is input and output sockets.
- **THREE-DIMENSIONAL (3D) VIDEO DISPLAYS:** each 3D pixel (voxel) of a cube-shaped space could be wired up with a separate wormhole, so that light fed in at the other end would appear in the voxel, creating a 3D image with invisible wiring.
- **MAGNETIC MONOPOLES:** the existence of magnetic particles that have only a 'north' or a 'south' pole has been long debated by physicists. But an artificial monopole could be made by conducting the magnetic field lines of just one pole of a magnet into a wormhole, so that the other end would act like a monopole.
- **MAGNETIC RESONANCE IMAGING (MRI):** a wormhole could transport the particles used for MRI, such as magnetic nanoparticles, to the imaged area without disturbing the applied magnetic field, which would allow high-quality images to be obtained.

P. B.

D. VAN RAVENSWAAY/SPL

Like space-time wormholes, electromagnetic wormholes connect remote regions of space.

have to be much smaller.

Greenleaf and his colleagues say that such a shield can be regarded as 'blowing up a point': expanding an infinitely small — and thus invisible — point in space while 'moulding' the way light interacts with it so as to sustain the invisibility. Their wormholes, in contrast, are like 'blowing up a line' — in essence, rather like making a tubular shield. They have worked out a prescription for the properties a metamaterial tube would need to have in order for light from outside to bend around it while light inside bounces along the channel as if along an optical fibre. In this way, light entering one end of the wormhole would be visible at the other end — albeit with some odd distortions. If the wormhole is very short, it acts rather like a fisheye lens, they say.

The researchers believe that it should be possible to make such devices with the same

microwave metamaterials as those used by Smith's team. But Leonhardt cautions that this remains "very far in the future". Making a 3D invisibility shield is already a big challenge, he says, and a wormhole would be harder still.

Smith is optimistic, however. "A version of these wormholes at microwave frequencies could definitely be feasible, although over a very narrow frequency range," he says. "I'm hoping to demonstrate lots of interesting optical structures using metamaterials in the near future, and we may add this one to the list." ■

**Philip Ball**

1. Greenleaf, A., Kurylev, Y., Lassas, M. & Uhlmann, G. *Phys. Rev. Lett.* **99**, 183901 (2007).
2. Schurig, D. *et al. Science* **314**, 977–980 (2006).
3. Leonhardt, U. *Science* **312**, 1777–1780 (2006).
4. Pendry, J. B., Schurig, D. & Smith, D. R. *Science* **312**, 1780–1782 (2006).

## ON THE RECORD

**"This is not one pig flying in orbit — this is a herd of pigs with gold trotters, platinum tails and diamond eyes."**

Gwyneth Dunwoody, chair of Britain's Transport Select Committee, chooses a striking metaphor to promote Europe's planned Galileo global-positioning system.

## SCORECARD



**Japanese wetland**  
Engineers are

painstakingly adding bends to the Kushiro River in Hokkaido to reverse the shrinking of the surrounding wetlands that has occurred since the waterway was straightened in the 1980s.



**Wet building**

The Massachusetts Institute of Technology is suing architect Frank Gehry over its Stata Center, which it claims is leaking and growing mouldy.

NAIC/ARECIBO OBSERVATORY/NSF



## 3 GOOD REASONS...

Why the Arecibo Observatory in Puerto Rico (pictured above) should not be shut down after 2011:

**1** It's the world's largest radio telescope — 20 times more sensitive than any other instrument for keeping tabs on near-Earth asteroids.

**2** It might save scientists from embarrassment — this week they accidentally identified the European satellite Rosetta as an 'incoming asteroid'.

**3** NASA seems to have the money to bail out the observatory, which needs to find US\$4 million per year after 2011 — coincidentally the same amount that NASA spent on parties this year.

Sources: BBC, planetary.org, CBS News, New York Times, Associated Press



## New Jersey rebuffs loan to fund stem-cell research

On 6 November, voters in New Jersey rejected a proposal to borrow \$450 million to fund stem-cell research. The move is the first time a bond measure has been defeated by voters in the state in 17 years. Pre-election polling had indicated that the measure would win easily, which made the 53% to 47% loss a startling rebuke to supporters.

Political observers suggest that fiscal concerns might have been a factor: New Jersey is \$30 billion in debt.

The vote came two weeks after construction work began on the state's first stem-cell institute. The 18-storey Stem Cell Institute of New Jersey in New Brunswick is being financed by \$270 million that state lawmakers approved in 2006. The \$450 million was to have funded research grants over ten years.

## Accelerator will bring antiproton beam online

A €1.2-billion (US\$1.7-billion) particle accelerator to be built in Germany has been given the final go-ahead. The Facility for Antiproton and Ion Research (FAIR) will investigate exotic atomic nuclei and examine the states of matter in the moments after the Big Bang.

The 15-nation consortium behind the project gave it the green light on 7 November. Construction is scheduled to begin in late 2008 and should be complete by 2016.

FAIR will feature a double-ring synchrotron with a circumference of 1,100 metres connected to the existing GSI facility in Darmstadt, Germany. It will be able to accelerate ions as heavy as uranium, and will also generate the world's most powerful antiproton beam.

Germany is picking up three-quarters of the price tag, the rest will be paid for by the partner countries, which include China, France, India, Italy, Russia and Britain.



The GSI's UNILAC will serve as a pre-accelerator for a planned double-ring synchrotron.

## Lunar pole captured on Moon video

Areas of the Moon's north polar region pop out in stark relief in this image taken by the Japanese lunar mission Kaguya (Selene).

The Japanese Aerospace Exploration Agency and the Japan Broadcasting Corporation, both based in Tokyo, teamed

up to broadcast high-definition television footage from Kaguya. This shot, taken on 31 October, is a still from the video of two short clips available at [http://www.jaxa.jp/topics/2007/11\\_e.html](http://www.jaxa.jp/topics/2007/11_e.html).

Taken from an elevation of 100 kilometres, it shows the landscape below the probe as it moves from the Oceanus Procellarum region towards the lunar north pole.



JAXA/NHK

## Knowledge gaps pour cold water on sea fertilization

The parties to the London Convention, the international treaty governing ocean dumping, have agreed that large-scale ocean 'fertilization' is currently not justified given gaps in scientific knowledge.

Adding iron, phosphate or urea can stimulate plankton growth in nutrient-poor parts of the ocean — thereby sucking down carbon dioxide from the atmosphere and potentially helping to combat climate change. But critics warn that the ecological impact is poorly understood, and at its meeting in London last week, the convention endorsed concerns issued in June by its scientific advisory group.

Environmentalists last week also raised concerns over an Australian company's plan to test its carbon-sequestration technology in Philippine waters. Ocean Nourishment Corporation (ONC) of Pyrmont, New South Wales, is currently in discussions with Philippine authorities over dumping 500 tonnes of dissolved urea, a nitrogen-based fertilizer, in the Sulu Sea between Borneo and the Philippines.

## Patient privacy rules hamper US research

US biomedical research is being hampered by privacy rules introduced in 2003, according to a survey of epidemiologists.

More than 1,500 'eligible professionals' took part in the survey, and 68% of them said that research has been made more difficult by the privacy rules (R. B. Ness *J. Am. Med. Assoc.* 298, 2164–2170; 2007). The rules are meant to ensure that medical records remain personal and private. But only 26% of the epidemiologists in the survey thought that the rules have enhanced confidentiality.

The survey, conducted by the Joint Policy Committee for the Societies of Epidemiology, gauged the effect of the Health Insurance Portability and Accountability Act Privacy Rule. This allows disclosure of patient records for research only if the scientist has authorization from the patient or has been granted a waiver by an institutional review board.

## Cap on overheads may put universities off defence

The US Congress has set a limit on overhead expenses for research done at universities on behalf of the Department of Defense.

The provision, approved in a final version of the defence-spending bill on 8 November, says that no more than 35% of research funding can be spent on costs such as maintenance and utilities. The bill now goes to President Bush to be signed into law.

"It's going to be a disincentive for the nation's best universities to conduct defence research," says Barry Toiv, a spokesman for the Association of American Universities in Washington DC. After the Department of Agriculture set a similar cap, he notes, fewer universities applied for government grants.

Currently, overhead expense rates are negotiated individually with the university department — the average is 34%, according to the latest survey from the Council on Governmental Relations.

### Correction

The photo that accompanied our News In Brief story 'Argo systems makes a splash with final float' (*Nature* 450, 148; 2007) should have been captioned as being a demonstration deployment, not the actual 3,000th float, and credited to Alan Blacklock, National Institute of Water and Atmospheric Research, New Zealand. In addition, the Argo data are accessible to all nations — not just the ones that participate in the project, which will continue to deploy additional floats over time to maintain the array.



# Political climate

American legislators are getting started on the first laws to tackle greenhouse-gas emissions. But Congress has a long way to go, says **David Goldston**.

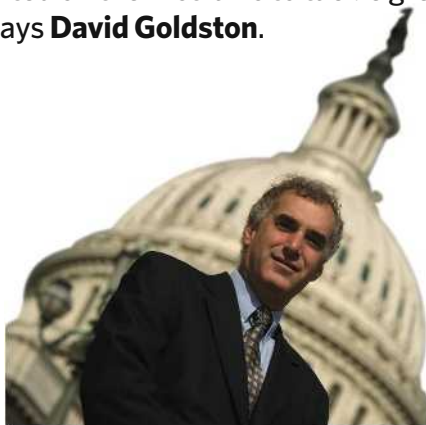
Is the US Congress ready to take action to limit global climate change? No. But a look at two key bills shows how much progress has been made since this time last year — and how many hurdles remain.

The legislation getting the most attention is a bill to mandate reductions in greenhouse-gas emissions, introduced by senators Joseph Lieberman (Independent, Connecticut) and John Warner (Republican, Virginia). They describe the measure as reducing emissions to as much as 10% below 2005 levels by 2020, and 63% by 2050.

On 1 November, a Senate subcommittee that Lieberman chairs approved the bill — just the first step in a tortuous legislative process, but the first time that any congressional panel has approved a measure that would require cuts in greenhouse-gas emissions. Moreover, the bill has gained support from middle-of-the-road members of both parties who have stayed away from previous climate measures, and a key swing senator from a coal state, Max Baucus (Democrat, Montana), voted for it in subcommittee.

That's a big difference from last year, when Lieberman backed a somewhat weaker bill he had introduced with John McCain (Republican, Arizona). With the Republicans in control back then, climate bills were not scheduled for committee votes — and they would not have passed in any event. But the Lieberman–Warner bill is not exactly sailing through Congress now; it squeezed by with a 4–3 vote in the panel most worried about climate change. Of even greater concern, the bill faced opposition not only from conservative Republicans, but also from the most liberal member of the panel, Bernie Sanders (Independent, Vermont). Echoing several environmental groups, Sanders argues that all emission credits in a cap-and-trade system should be auctioned off, whereas the bill provides some credits to corporations for free.

In one sense, the attacks from the left are a good sign. Proponents of action used to be happy to see almost any measure that acknowledged the need to cut emissions; nothing was going to pass anyway, so just getting officials on-the-record could be counted as progress. But now the details matter, and Lieberman and Warner are going to have to negotiate with both the right and the left. From the centre, they will also have



## PARTY OF ONE

to contend with a more restrained alternative offered by senators Jeff Bingaman (Democrat, New Mexico) and Arlen Specter (Republican, Pennsylvania). Cobbling together the votes for Senate passage has to be seen as a long-shot goal that is not likely to be met in this Congress.

If a bill ever did pass the Senate, it would probably face even greater obstacles in the House of Representatives, a more ideological body with a keener tendency to protect local economic interests. House Speaker Nancy Pelosi (Democrat, California) surprised everyone earlier this year by announcing that climate legislation would be a priority in the House and by setting up a new committee on the issue. But since then, climate bills seem to have slipped gradually from the House agenda, replaced, in part, by a related energy measure.

Different versions of the energy bill passed the House and Senate this summer, and both are largely a hodge-podge of less controversial ideas. But each version included at least one lightning rod — in the Senate, tighter mileage standards for cars, and in the House, renewable-energy requirements for utilities. If the House and Senate can't agree on those two items — and negotiations have been moving slowly — then it's hard to imagine action on a comprehensive climate measure, which would necessarily include much more controversial provisions. Indeed, some environmental groups view the bill as an unwelcome distraction, eating up the time of officials and staff who could be working on broader legislation. Worse still, they fear that if the energy bill passes, the transport and utility industries will claim that they already have been forced to make sacrifices and should therefore be treated gingerly in a larger climate bill.

'If' is a key word; the energy negotiations are clearly no cakewalk. The Senate approved the tougher mileage standards without a roll-call vote — in which all senators go on-the-record with their stance — so it's not clear how much latent opposition to the idea remains in the Senate, which had always defeated the standards by a sizeable margin in the past. The House, which has also defeated mileage standards in past years, didn't even take up the issue because its Democratic leadership asked proponents to hold off. Either the leadership didn't want the House on record defeating the standards, or it didn't want to put Democrats in the uncomfortable position of choosing between their environmental backers and their labour backers, who generally oppose the standards. Either explanation signals rough sledding for the proposal. The future of the utility provision is equally problematic.

Many observers think that an energy bill will finally emerge, in part because Pelosi has so vocally promised to produce one. The political fall-out from such a bill could be one factor that determines how much stomach Congress has for a fight over a measure that targets climate change. If Democrats are rewarded by the public and the media for taking long-debated steps to reduce fossil-fuel consumption, then that may encourage more high-profile action on the climate. It might also make the issue a more central part of the US presidential campaign.

But high-profile action is not necessarily speedy action. Complicated and controversial legislation usually takes several Congresses to pass. The first clean-air law was enacted in 1955, but no measure that had any real impact passed until the 1970s, and a real crackdown on many pollutants didn't begin until the 1990 version of the Clean Air Act. Partly, this is because no one can foresee all the intellectual and political issues that a bill raises until the detailed writing of it begins — and sometimes problems can't be foreseen until it is implemented. Congress has wasted years on an ideological debate on the reality of climate change; now at least it's getting down to the nitty-gritty. The biggest hurdle of all facing climate change legislation may be the steepness of the learning curve; Congress won't be able to make up all that lost time. ■

**David Goldston** is a visiting lecturer at Princeton University's Woodrow Wilson School of Public and International Affairs.

## BUSINESS

# Worth its weight in platinum

Booming mineral prices leave car makers scrambling to eke more catalytic performance out of precious metals. **Jeff Tollefson** reports.

Lovers may adore platinum's silvery sheen in their favourite jewellery — but at least they can switch to gold or silver if the price isn't right. Chemical engineers wish they had the same option.

Thanks to their role in catalytic converters, platinum, palladium and rhodium have all become crucial bulwarks in the fight against air pollution. Platinum and palladium can catalyse reactions that convert hydrocarbons and carbon monoxide into carbon dioxide and water vapour. Rhodium catalyses another reaction, converting nitrogen oxides into nitrogen and oxygen. These reactions are at the heart of the three-way catalytic converters now fitted on almost all cars in the developed world and on an increasing share of those in India, China and elsewhere.

Although materials such as gold and nickel can perform similar tricks at lower temperatures, only these platinum-group metals can do the job at the high pressures and temperatures of 900 °C or more that are found in vehicle exhaust systems. With demand from the global automobile industry pushing prices ever higher and no alternative catalysts on the immediate horizon, car manufacturers are now locked in a race to make each ingot go that a little bit further.

"From a scientific standpoint, the three-way catalytic converter is really a done deal. It's now an engineering feat to bring the volume of metals down," says Johannes Schwank, a chemical engineer at the University of Michigan in Ann Arbor.

Chemical engineers are working on the problem at the molecular level, trying to make the catalyst layers even thinner and diluting them with cheaper alloys. Equally important is the design of the underlying structure to which the catalysts are applied — although the location of the catalytic converter and engine design can also play parts. The competition is taking place largely beyond the academic eye in industrial labs around the world.

Catalytic converters were first put into use in the 1970s after the world's first regulations on car emissions came into effect in California. Automobile firms in the United States developed the technology based on advances in surface chemistry made during the previous decade.

General Motors, Ford and Chrysler were all involved in the development of the catalytic converter, which Schwank regards as one of the greatest achievements in the history



Catalytic-converter production in Shanghai: global demand for the units is surging.

of chemical engineering. He compares the hydrocarbon, carbon monoxide and nitrogen oxide molecules in the converters to helicopters trying to land on the roof of a hospital; the electrochemical properties of platinum, palladium and rhodium provide an ideal landing pad. Other elements might not allow them to land at all, or might "hold them hostage" after their arrival, he says. "The strength of the bonds is just about right to carry out the reactions, but not so strong that the reaction agents won't have a chance to leave."

At first, car makers relied on platinum for this function, but eventually switched to its sister metal, palladium, at one-third of the price. When palladium prices spiked in 2000, they went back to platinum, but the cycle has now repeated itself. Platinum prices are approaching US\$1,500 per troy ounce (this historic measure of precious metals is the equivalent of about 31 grams); palladium costs less than \$400.

The recent march of car makers back to palladium has helped to moderate the price of platinum, as has slower demand from jewelers, according to Johnson Matthey, a leading catalyst-making firm based in London. However, platinum performs better than palladium in diesel exhaust systems.

Global platinum supplies stood at more than 7.6 million troy ounces in 2006, 66% of which was used in autocatalysts, according to Johnson Matthey. Those figures include the recovery of 855,000 troy ounces of platinum from recycled autocatalysts, which provided 11% of overall global demand. South Africa supplied 78% of the world's new platinum in 2006 and is home to the only mines in which platinum is a primary product. For the decade ending in 2006, platinum production increased by more than one-third — but demand for autocatalysts more than doubled.

Platinum and palladium can play off each other, but manufacturers have to pay whatever the market demands for rhodium, for which there is no alternative. Posted at around \$6,500 per troy ounce — almost eight times the price of gold — rhodium is one of the most expensive elements on Earth. A staggering 87% of the world supply goes into autocatalysts.

Jeremy Coombes, an analyst for Johnson Matthey, says that the amounts of the metals used in each catalyst depend on the size of the car, the kind of fuel being used and the local air regulations, and can range from 1–2 grams for

JOHNSON MATTHEY/PLATINUM 2004

a small car in a lightly regulated environment to 12–15 grams for a big truck in the United States. That translates to anywhere from \$25 to a few hundred dollars per vehicle, he says: a significant amount for the likes of Toyota, which sold 2.5 million vehicles last year.

Autocatalysts often use honeycomb-like structures to create vast surface areas — equivalent to perhaps a couple of football pitches in each converter — in which reactions can take place. The catalysts are layered as thinly as possible, often by dipping a ceramic structure into a solution of the metal. The gradual agglomeration of metal particles under intense heat, eventually reducing the surface area and catalyst efficiency, is one of catalyst designers' main challenges.

Nissan announced in August that it has deployed a new fabrication method that uses nanotechnology to reduce agglomeration. The company claimed the technique would halve the use of precious metals in its catalytic converters. Mazda said last month that it can now embed nanoparticles made of platinum-group metals into ceramic spheres, cutting use "by 70–90% with the same level of purifying efficiency". But no public data are available to verify these claims.

Although few see prospects for replacing platinum-group metals in catalytic converters, researchers are searching hard for alternatives in applications that use low-temperature catalysis. Designers of fuel cells, which catalyse hydrogen and oxygen to produce electricity and water vapour, are eagerly seeking cheaper alternatives to platinum. And Schwank says that nickel has shown promise as a catalyst for onboard fuel reforming, a technique for generating hydrogen from fossil fuels for use in fuel cells.

And although advances in catalyst design are likely to further reduce the amount of platinum, palladium or rhodium needed in each vehicle, those gains might not be enough. Global automobile production has risen from 56 million vehicles in 2000 to more than 66 million last year and it shows no signs of slowing, according to the consulting firm Global Insight in Waltham, Massachusetts.

"There is simply not enough platinum and rhodium going round on this planet to satisfy the collective demand of automotive emission-control systems and all of these other areas," Schwank says. Supply is sure to follow demand upwards in years to come, but few analysts expect it to get out in front, and pull prices down. That will leave new technologies that need the platinum-group metals, such as fuel cells, paying a heavy price for autocatalysts' insatiable appetite for them. ■

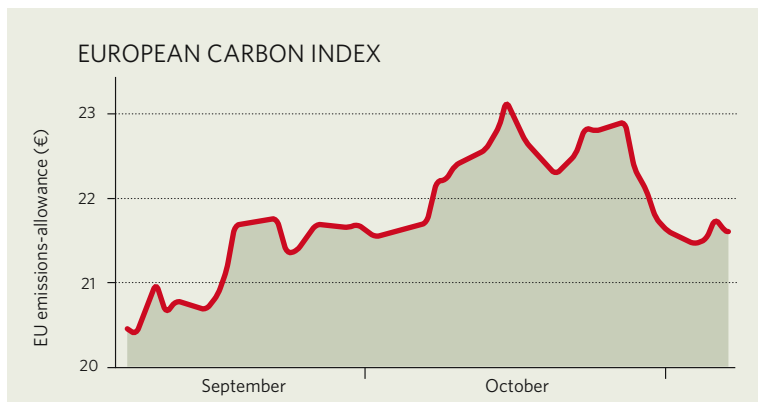
## IN BRIEF

**PULLBACK FROM PROTEINS** Genomics company Celera in Rockville, Maryland, has confirmed that it has laid off as many as half of its 50-odd proteomics researchers over the past six months. "The business is shifting from discovery into development, and that's throughout the organization," says David Shipley, a spokesman for the company, who declined to specify the exact number of departures. He added that the cuts at the company, which employs 540 people, were confined to proteomics. The announcement came as Celera reported its first-ever profitable quarter, and on the heels of its October acquisitions of two firms in the San Francisco Bay area — Berkeley HeartLab and Atria Genetics.

**TAG TEAM** Ford and Daimler joined forces to buy the automotive fuel-cell division of Ballard Power Systems of British Columbia, Canada, in a deal worth \$228 million. Ballard is one of the leading developers of fuel cells, which burn hydrogen and remain a long way from commercialization: some analysts estimate that costs need to come down a hundred-fold to make them competitive with internal-combustion engines. But the move by the two big rival car companies was seen by motor-industry analysts as a significant step toward the eventual adaptation of fuel cells into the mainstream motor industry.

**FLU DRUG FALL** Chugai, the Japanese distributor of Tamiflu, has said that it will halve the amount of the drug it is preparing to sell this winter in response to the precipitous drop in demand, brought about by reports earlier this year associating it with teenage suicide (see *Nature* 446, 358–359 and 481; 2007). In March, the government issued a safety warning against prescribing the drug for teenagers. The fall is a serious blow to Tokyo-based Chugai and to Roche, its parent company based in Basel, Switzerland, which makes the drug. Japan has in the past accounted for up to 70% of Tamiflu's global sales.

## MARKET WATCH



The price of carbon allowances for the European Union's emissions-trading scheme for 2008 to 2012 increased moderately in late summer and early autumn, but has dropped slightly since mid-October. The trend is being driven, analysts say, by rising prices for oil and natural gas, which makes it profitable for energy utilities to switch to coal, which is cheaper but more carbon dioxide-intensive.

A 'carbon credit' to emit one extra tonne of CO<sub>2</sub> during 2008, for example, last week cost €22 (US\$32) at the European Energy Exchange in Leipzig, Germany — up from €18 in late-August.

Price volatility for the allowances is likely to become more pronounced in the winter, market watchers say, when trade volumes at Europe's five carbon exchanges tend to peak. And some of them expect the price to rise further.

"A price of €25–30 is not unrealistic in the near future," says Stefan Kleeberg, a carbon-market analyst with the 3C Group near Frankfurt, Germany.

Prices on the carbon-credit market could also rise as a result of delays in the certification of clean-development projects in poorer nations, says Kleeberg. Shares in the leading British carbon-trading specialist EcoSecurities slumped by almost half last week after the company announced that some 30 million tonnes of carbon credits that had been anticipated by 2012 would probably not materialize.

The price trend comes against a backdrop of continued growth in global carbon emissions, according to the International Energy Agency's latest World Energy Outlook, released on 7 November.

**Quirin Schiermeier** ■





# SHOWDOWN IN A SUNBURNT COUNTRY

Climate is shaping up as an issue in the 24 November Australian elections, as **Stephen Pincock** reports.

**A**ny campaign veteran will tell you that voters are fickle, switching from candidate to candidate and issue to issue as the whim takes them. But in Australia, voters may have changed their minds once and for all on the issue of climate change. In mid-2006, something seemed to shift climate from an 'issue of concern' to the top of the list of people's most serious considerations.

Frank Muller, an expert on sustainability policy at the University of New South Wales in Sydney, recalls that the change took place while he was on a four-month visit to America last year. "I went to the United States in June and was back in October, and in that period, this big switch had taken place," he says. Others pinpoint it even more precisely. "I think in about September of last year there was a global shift in awareness of this issue," says Tim Flannery, a zoologist at Macquarie University in Sydney who won this year's Australian of the Year award for his climate activism.

Blame any number of factors for the switch: Al Gore's visit to Australia in September 2006 (and Prime Minister John Howard's refusal to meet with him); the October 2006 release of Nicholas Stern's review of the economics of climate change, which estimated vast costs if global warming is not stemmed soon; and,

looming over all, the drought that is plaguing Australia, by some measures the worst in a century or more.

The timing could not be worse for Howard (pictured above), the Liberal prime minister who has led Australia for 11 years and is behind in opinion polls in his quest for a fifth term in office. On 24 November, Howard will face his Labor opponent, Kevin Rudd, in a federal election to determine who will form the next national government. And climate is shaping up as a major election issue. One of Howard's defining foreign-policy stances has been his refusal to ratify the Kyoto Protocol to control greenhouse-gas emissions; Rudd has vowed to ratify it immediately if elected.

As the election date draws closer, both parties have released a slew of climate-related policies and promises in an effort to woo voters. They range from Rudd's Aus\$200-million (US\$185-million) plan to save the Great Barrier Reef, to Howard's vow to ratify an international agreement to replace Kyoto after it expires in 2012 — that is, as long as the agreement applies to major emitters such as China and India.

What happens in Australia's election could

shape the international dynamics on climate change for years to come. A week afterwards, representatives will gather in neighbouring Bali to start thrashing out details of a post-Kyoto strategy for limiting emissions (see page 319). A change in Australian leadership could bolster morale, if not much else, among the delegates. And with a presidential election

due the following November in the United States (see page 340), some envisage a not-so-distant future in which the leaders of the two major countries that did not ratify Kyoto are instead in favour of mandatory reductions in greenhouse gases.

For Australians, the potential consequences of climate change have been driven home by the country's water crisis. The five-year drought, often described as the worst in living memory, has left rural communities in the southeast of the country reeling and many urban areas with severe water shortages. Average annual inflows into Sydney's dams for the period 1991–2006, for instance, were 71% less than for 1948–1990.

The drought is hitting home in many parts of the country, says Mike Young, professor of water economics and management at the

**"In their minds, Australians are linking the drought to the longer-term issue of climate change."**

I. WALDIE/GETTY IMAGES



University of Adelaide. "People are looking at water shortages," he says, "and farmers are finding that their water entitlements have been dramatically debased." Young says that in their minds, Australians are linking the drought to the longer-term issue of climate change.

Opinion polls bear that out. In August, a poll conducted by the Lowy Institute for International Policy, a Sydney-based think-tank, suggested that climate change ranks ahead of nuclear weapons, Islamic fundamentalism and international terrorism as the external threat most Australians are concerned about. And earlier this month, the Climate Institute in Sydney commissioned a poll of 877 voters in 9 key marginal electorates. It found that 73% of voters thought climate change would have either a strong or a very strong influence on their vote at the election, an increase from 62% in August.

### All change

Howard's anti-Kyoto stance did not prevent him from being re-elected in 1998, 2001 and 2004. But that may be about to change. Although Australians aren't likely to cast their vote solely on the basis of climate-change policies, the divergent positions of Howard and Rudd are becoming a major force in the current election campaign, says veteran pollster David Briggs from Galaxy Research in Sydney.

"It is one of those areas where Rudd has been able to extract some point of difference," says Briggs. "It does make a contribution to the perception that Rudd is the man of the future, and that he has a vision for the future that includes proper environmental planning, whereas John Howard is a man of the past."

For much of his time in office, Howard has questioned the scientific basis of anthropogenic climate change, says Peter van Onselen, a political scientist at Edith Cowan University in Perth and co-author of a biography on the prime minister. "He was comfortably within the category of being a climate-change sceptic," van Onselen says. "He felt that the science on it, at least what was presented to him, went both ways as to the significance of it or not."

Guy Pearce, a former speechwriter for Robert Hill, one of Howard's previous environment ministers, argues that the prime minister's perspective was heavily influenced by Australia's mining and energy sectors. Pearce's book *High & Dry* details, on the basis of interviews with key players in the country's greenhouse-gas policies,



Kevin Rudd wants to ratify the Kyoto Protocol.

what he calls the 'greenhouse mafia'. According to Pearce, lobbyists and advisers from industry groups sold Howard the idea that Australia's economy rests on the supply of cheap fossil fuels. Their message, he says, was simple: "When it all boils down, it's about avoiding cuts in Australia's emissions for as long as possible, delaying as long as possible."

But even within Howard's cabinet, there has been dissent on the subject of climate. In late October, the *Australian Financial Review* broke a story that Howard's environment minister, Malcolm Turnbull, had six weeks earlier failed in his attempt to convince the cabinet to ratify the Kyoto agreement. Howard has not denied the charge, and Turnbull has not spoken publicly about it.

Analysts agree that Howard tends to view

the issue of climate change through the prism of economics. Like US President George W. Bush, Howard has long said that he thinks mandatory emissions reductions could hurt his country's economic growth, and that excluding developing countries such as China from the agreement would hurt Australia competitively. This came to the forefront in 2002, when Howard decided not to ratify the Kyoto Protocol, despite having signed it.

### Clean green

In place of mandatory emissions reductions, Howard opted for several clean-energy initiatives. In June 2004, for instance, his Liberal-National coalition government announced a new policy called Securing Australia's Energy Future. It includes a Aus\$500-million fund to encourage the energy sector to develop lower-emission technologies, and Aus\$75 million targeted at solar-energy projects. Four months later, Howard was re-elected.

Environmental groups have criticized Howard's continuing focus on fossil fuels. According to the Australian Bureau of Agricultural and Resource Economics, in 2004–05, 93% of the country's electricity was generated from fossil fuels (coal, oil and gas), and 7% from renewables such as hydroelectricity, wind, biomass and biogas. That could have something to do with the nation's 73-billion-tonne coal deposits, found mostly in New South Wales and Queensland. Coal also forms the backbone of a robust trade with nearby China. "Whatever may be the merits of renewables," Howard told ABC radio in 2004, "the reality is that the older fuels of which we have large supplies are going to contribute the bulk of our energy needs and what we have to do is to make them cleaner." Hence his government's focus on emphasizing clean-coal technologies, such as carbon capture and sequestration, and on nuclear power as an alternative low-emission energy source in the future.

The Howard government has made several major, if criticized, initiatives in the field of climate. In 2005, Howard took Australia into the Asia-Pacific Partnership on Clean Development and Climate, a non-binding partnership established with China, India, Japan, South Korea and the United States to foster cooperation on climate-change action. Australia pledged Aus\$100 million to the group over five years, from which it has so far allocated Aus\$60 million to 44 projects. These include deployment of high-efficiency solar-power stations, improved



Water restrictions have brought the issue of climate change to the fore.

M. NOLAN/GETTY IMAGES

L. WALDIE/GETTY IMAGES



Australia's wealth of coal has made developing alternative energy sources a relatively low priority.

efficiency standards for appliances and a mobile system that can test carbon capture at coal-fired power stations.

In June 2007, Howard announced that the government would introduce a cap-and-trade emissions-trading scheme, which observers consider his most significant backflip on climate-change policy in the lead-up to the election.

And in September, Howard placed climate change on the agenda for a meeting of the Asia-Pacific Economic Cooperation (APEC) forum, which includes the United States, China, Russia and Japan. Without setting any hard targets, the APEC leaders signed up to a statement, which said they agreed "to work to achieve a common understanding on a long-term aspirational global emissions reduction goal to pave the way for an effective post-2012 international arrangement".

### Conservative approach

The question for many observers is whether such actions will be enough to make a difference. In the run-up to the elections, Australian climate scientists are cautious about seeming overly partisan, and even the outspoken Flannery is reticent about openly criticizing any political party. But if pressed, he notes that Howard's climate announcements lack teeth. "If you add up all of those policies, all of the pronouncements," he says, "they make no difference in terms of emissions."

In 2005, greenhouse-gas emissions in Australia reached 559 million tonnes of carbon dioxide equivalent, which is 2% higher than 1990 levels. Government figures show that per capita emissions fell between 1990 and 2005, from 32.3 to 27.6 tonnes of carbon

dioxide equivalent — but they remain the second highest in the developed world, after Luxembourg.

Dave Griggs, a climate scientist at Monash University in Melbourne and former head of the science working group secretariat of the Intergovernmental Panel on Climate Change (IPCC), argues that Australia's position in the ranks is not acceptable. "The IPCC fourth assessment report is very clear that we need to mitigate greenhouse-gas emissions by a very large fraction," he says. "I don't see that the policies that have been put in place to date have been aiming to get an Australian economy in which emissions have gone down to a level that is probably required globally. Have current policies put us on the right track? The answer is clearly no."

Climate is, of course, just one issue facing voters in the upcoming elections; the Howard and Rudd campaigns have also included ongoing concerns such as the economy, health care and taxation. For some, the prominence of climate echoes previous elections in which environmental issues played a role. In 1983 for example, a controversial proposal to dam a Tasmanian river helped to bring down the government of Malcolm Fraser. In the upcoming election, "the difference between the major parties over whether or not to ratify the Kyoto Protocol is as sharp as the difference in 1983 over the Franklin dam," says Muller.

Before the election was called, for example, the Howard government committed to ensuring that about 15% of Australia's electricity

would come from low-emission sources by 2020. It also launched a taxpayer-funded advertising campaign, featuring television commercials that focused on simple things families could do to reduce their greenhouse footprint, such as drying clothes on a washing line or switching to compact fluorescent lights (incandescent bulbs are being phased out entirely across the country).

### Power and the passion

Both parties have also opted for charismatic, high-profile figures as their environment representatives. In the case of the Labor party, the candidate is Peter Garrett, who was the front man for the rock group Midnight Oil and former president of the advocacy group Australian Conservation Foundation in Melbourne. The government's choice for environment minister is Turnbull, the minister who reportedly argued for ratifying Kyoto.

Howard and Rudd themselves have focused on climate change at key moments. During the only televised debate between them in the campaign, on 21 October, Howard chose climate change for one of two new policy announcements (the other being on troops in Iraq). He vowed to establish a fund to pay for development of clean energy and to compensate low-income earners for hikes in power costs, funded by the revenue raised by auctioning emissions permits.

For his part, Rudd made a point of reminding voters that his party had set a concrete target for a 60% reduction in emissions from 2000 levels by 2050. The Labor party has, however, refused

to commit to a target for 2020 until after it receives an economic report it commissioned on climate change from Ross Garnaut from the Australian National University in Canberra.

Last month, the Australian Conservation Foundation released an election scorecard designed to compare the parties

on their approaches to climate change, sustainability and the environment. With four weeks to go, both parties scored poorly: Labor at 56%; Howard's coalition just 21%.

All this leads some commentators to despair about whether the election can truly turn around the major climate issues facing Australia. "We love to romanticize about a sunburnt country," says Pearse, referring to the well-loved Australian poem *My Country*. "But people need to start realizing that what we're talking about is a country burnt beyond recognition in our lifetimes — and unless we start acting, that's what we're going to cop." ■

Stephen Pincock is a writer based in Sydney.

**"Have current policies put us on the right track? The answer is clearly no."**

— Dave Griggs





Rudy Giuliani and Hillary Clinton could bring very different approaches to addressing climate change if elected president.

# Beyond Bush

The next US president could lead the country into meaningful action on controlling greenhouse-gas emissions, but only if he, or she, can seize the moment. **Jeff Tollefson** reports.

It is 20 January 2009, and the new US president has just been sworn into the Oval Office. Environmentalists are optimistic. Scientists, activists and politicians around the globe have their fingers crossed. For the first time in eight years, the United States is led by a politician who advocates quick and forceful action on global warming.

The question on everybody's mind is what comes next. It doesn't take much to prepare a policy paper and offer campaign sound bites about the dangers of global warming. It's also easy to grandstand one's green credentials once in office — perhaps by requiring Cabinet members to offset their carbon footprints or by departing the inauguration in a hybrid vehicle (although the armour-plating might well hamper its gas mileage). But it is a lot harder to bind the world's leading economy to a meaningful regulatory regime for greenhouse-gas emissions.

"We are talking about the largest environmental policy we have ever implemented, in terms of economic costs and economic benefits," says Joe Aldy, a fellow at Resources for the Future, a non-partisan think-tank in Washington DC. "When you talk about policies of that stature, you have to have executive leadership."

So what can the new president actually accomplish in those first days? The most immediate changes could be seen in the fed-

eral budget — in which billions of dollars are in play in research money for energy and global warming — or in new regulations controlling energy efficiency or carbon-dioxide emissions. Over the course of a presidency, though, the measure of true action would come in shepherding climate legislation through the halls of Congress, while leading the world in developing a successor to the Kyoto Protocol to control greenhouse gases.

On the domestic front, the Supreme Court set the stage for at least some new regulations in April, when it ruled that the Environmental Protection Agency (EPA) has the authority to regulate carbon dioxide from vehicles (see *Nature* **446**, 589; 2007).

Regulating the gas would represent an old-school approach to environmental issues that forces every business to meet the same standards. This system has fallen out of favour with the advent of cap-and-trade systems in which businesses barter among themselves for the cheapest way to meet an overall mandate, but a president could use the threat of regulation to force action on meaningful cap-and-trade legislation in Congress. "It's really important to recognize," says Aldy, "that the alternative to a new climate bill coming out of Congress is not no carbon regulation, which is the status quo, but very costly carbon regulation."

A new president could even interpret the Supreme Court ruling to include carbon-dioxide

emissions from power plants. "If you had decent regulatory standards for vehicles, fuels and power plants, you will have dealt with 60% of US emissions," says David Doniger, climate-policy director for the Natural Resources Defense Council, an environmental group based in New York.

## Capitol support

The next administration is likely to face a Congress in which both houses support some sort of action on global warming. Several major emissions-reduction bills have already been introduced (see pages 333 and 342), but most experts expect the debate to extend well into the next administration. And while congressional leaders set the legislative agenda, the president can set the tone on whether the executive branch will back up any particular action. The White House can, for instance, tap a team of experts to meet with industry and activist groups, shuffle proposals to and from Capitol Hill, and otherwise ensure that the issue receives the sustained attention necessary to iron out an agreement.

When a president is willing, practical politics can sometimes follow swiftly. On coming into office in 1989, for example, President George H. W. Bush made action on acid rain a priority. Top presidential advisers on domestic policy, the environment and economics began working on the issue, says Jeff Holmstead, who joined the team in 1989 and went on to head the air and radiation division at the EPA under the current administration. "For them it was

**"We are talking about the largest environmental policy we have ever implemented."**

— Joe Aldy

C. SENTER/AP; C. DHARAPAK/AP

essentially a full-time job," he says. "These were the senior-most people in the White House who would meet about this every day." As a result, legislation that had floundered in Congress for nearly a decade took off, and amendments to the Clean Air Act passed the following year, establishing a cap-and-trade programme for sulphur dioxide that would later serve as a model for the Kyoto Protocol.

A president could also take advantage of manpower and expertise that Congress lacks. In this case, that means marshalling career experts at agencies such as the EPA and the Department of Energy to tackle global warming. Currently in climate legislation, says Doniger, "you have literally a handful of staff on the Hill trying to put the whole thing together, and there are a lot of experts in the agencies who are being held back. If you are trying to build a skyscraper and you are trying to learn welding at the same time, it's more difficult than if you have all the welders there with you".

### Bureaucratic challenges

Another immediate move that a new president could make would be to restructure the government's \$1.7-billion Climate Change Science Program, which coordinates climate research among 13 agencies. A recent National Academies report criticized the red tape ensnaring the programme, which reports to multiple organizations within the White House. "This is such a huge challenge that you can't have it mired in bureaucracy," says Eileen Claussen, president of the Pew Center on Global Climate Change in Arlington, Virginia. "You have to have direct links into the decision makers, and that means the cabinet level and the White House."

The administration could also make it easier for individual states to pursue their own programmes. The administration of George W. Bush has so far refused to rule on a petition from California that seeks permission to regulate carbon-dioxide emissions from automobiles. Even if Bush decides to deny that petition in favour of the federal regulations he

is expected to issue next year, a new president might be able to revisit that.

Getting the domestic agenda in order is just the first major step for a new president. The next would be coordinating that with a major international change in agenda, which is something only a president can do. "You've got to work on parallel tracks," says Dan Esty, director of the Yale Center for Environmental Law and Policy in New Haven, Connecticut. "You don't bring a bill to Congress until you've got an international treaty tied down, but you don't go to the international community without a series of domestic packages."

President Bush has alienated much of the world by refusing to endorse any mandatory greenhouse-gas reductions, but the new administration will be in a position to assume a leadership role as the international community works toward a post-Kyoto agreement. "All the president has to do is essentially get on the phone to the European Union, the G8 countries, the other major economies and developing nations, and begin talking," says Robert Stavins, an economist at Harvard University in Cambridge, Massachusetts.

Some experts believe the new president might even be able to push a new approach onto the international agenda if fellow world leaders believe the United States is serious about global warming. Yvo de Boer, executive secretary of the United Nations Framework Convention on Climate Change, has already signalled a willingness to entertain alternative ideas to ensure that everybody is on board the next global climate treaty. He says the notion of using a variety of national commitments — things such as renewable-energy standards, fuel-efficiency regulations or even elimination of energy subsidies — might be worth considering in addition to Kyoto-style carbon caps. Different types of commitments would be available to countries facing various economic and political realities.

De Boer also downplays his expectations of progress under a new administration, pointing out that Kyoto itself serves as a reminder of the limitations of the US presidency and the power of lawmakers on Capitol Hill. "It was President Clinton who never took the Kyoto Protocol to the Senate, because he knew he wouldn't get it through," he says. "The Kyoto experience shows that if an administration negotiates something that the Senate isn't willing to ratify, you are not really that much better off."

### Internal doubts

As if leading the free world weren't enough, the new president will also have to address

any lingering scepticism from the American public. Global warming consistently ranks as an issue of concern to US voters, but some polls put it far behind other national priorities such as terrorism, education, immigration and health care. A new leader might even be able to help activists establish a new, more hopeful icon for global warming. Polar bears on melting ice floes might not be enough to convince Americans, long accustomed to cheap energy, that \$5-per-gallon gasoline is acceptable, fears Severin Borenstein, an economist at the University of California, Berkeley. "But maybe a charismatic leader can get people to make sacrifices over the long term," he says.

In the end, many experts expect that whatever the next president achieves domestically will set the standard for a post-Kyoto agreement. Many international delegates hope to achieve an agreement on the post-Kyoto framework by 2009, an aggressive goal that leaves the new president less than a year to get everything done. "It's more realistic to look at early-to-mid 2010," says Esty. "Three or four months for the administration to get its feet on the ground, a year to get the negotiations done. It's a fast pace, but a doable one."

**Jeff Tollefson covers climate, energy and the environment for *Nature*.**

**"All the president has to do is essentially get on the phone and begin talking."**

— Robert Stavins

J. L. MAGANA/AP; S. SENNE/AP

J. RAOUX/AP; C. NEIBERGALL/AP



Charismatic candidates, but how green are their policies? From left to right: Fred Thompson, Barack Obama, John McCain and John Edwards.





M. TEMCHINE

Striking a balance: from left, Jonathan Pershing, Julian Braithwaite and Jason Grumet.

# THE FIRST CUT

For the first time, the US Congress has begun crafting comprehensive legislation to tackle global warming. *Nature* brought together five experts with various backgrounds to discuss the current political climate as the United States moves towards mandatory emissions caps.

**Two leading climate bills are currently before the Senate. One, by senators John Warner (Republican, Virginia) and Joseph Lieberman (Independent, Connecticut), would see a 63% reduction in emissions by 2050. The other, by Jeff Bingaman (Democrat, New Mexico) and Arlen Specter (Republican, Pennsylvania), offers softer targets. What are the advantages and disadvantages of the two?**

**Jason Grumet:** The fundamental architecture of the bills is extremely similar. I think that's what gives me some real optimism — that we can now see legislation because we have two serious, detailed, bipartisan proposals that have a great deal in common. There are some important differences, but I think it's worth noting that those differences are not philosophical or ideological. They are different approaches to trying to achieve similar ends.

**Jim Rogers:** It is important to try to minimize any disproportionate or adverse impact on any

certain region of the country because of historic decisions that were made about the type of fuel they use to generate electricity. So that debate will be ongoing. It's a question of fairness, so there isn't a formula that necessarily works for everybody. And at the end of the day, it's going to create a certain amount of pain — like all difficult compromises — for everybody, because there is no perfect solution.

**What are the chances we can get this done this autumn or in this Congress?**

**Grumet:** It is unlikely that we will see legislation signed by the president this year. However, I think there is a real possibility that we can see legislation garner 60 votes in the Senate if there is a political will to do so. And once that happens, that legislation becomes the centre of gravity for what will ultimately pass the Congress. There is a tactical question that I think our commission [the National Commission on Energy Policy] has some

disagreement with other advocates about, and that is whether in fact the legislation could be so much stronger environmentally in 2009 or 2010 that the best thing to do ecologically would be to actually not try to pass legislation in this Congress. Our commission fundamentally believes the opposite, that the most ecologically responsible thing to do is to get an economy-wide mandatory programme adopted in the United States. Our view is that we should get on with it.

**Rogers:** Every major piece of environmental legislation that has ever been adopted in this country has been overwhelmingly adopted bipartisanship, and it's never been a close vote. It's been huge votes. This goes all the way back to the Clean Air Act in 1970. So what I hope I will see in the future is this strong consensus growing out of both houses [of Congress].

**Jonathan Pershing:** It does seem to me that another piece that Jim didn't mention, but I think is quite critical, is the level of popular



## Who's who

**Julian Braithwaite** works on climate issues as counsellor for global issues at the British Embassy in Washington DC.

**Jason Grumet** is executive director of the National Commission on Energy Policy, a non-partisan group whose work served as a platform for Senator Bingaman's legislation.

**Michael MacCracken** is the chief scientist for the Climate Institute in Washington DC.

**Jonathan Pershing** heads the climate, energy and pollution programme for the World Resources Institute in Washington DC.

**Jim Rogers** is the chairman and chief executive of one of the nation's largest electric utilities, Duke Energy in Charlotte, North Carolina.

support. My own sense is that if I look back over the past three to five years, I've seen an increase in the willingness to pay [for action on global warming]. That brings us back to the question of what kind of technology costs there are, what kind of options we have and what kind of price we might need to pay to get to those options.

**Grumet:** The Bingaman–Specter bill has incentives that, based on permanent prices, would cost somewhere between \$35 and \$50 per tonne of carbon in the very first year of the programme for carbon capture and sequestration. Because any carbon price generates a significant amount of revenue, one of the key aspects of this legislation when it ultimately passes is going to be those technology incentives.

**Rogers:** I see this really through the eyes of our customers and what the cost impacts will be. It is fundamental in my judgement, and we can and should and will build a bridge to a low-carbon world. To build that bridge, we are going to have to build it on technology. [Developing that technology might take] 10–15 years for carbon capture, 10 years for renewables and storage, and 10 years for the nuclear option, at a minimum. We almost have to have a timeframe of how long it is going to take to get the technology that allows the bridge to be built.

**Pershing:** If we are looking at a process in which we wait for 10 to 15 years before we have any significant movement on a technology, particularly in the case of capture and storage, we've lost a substantial part of the battle. How do we get it so it's not a 10–15 year pathway, but a 5–8 year pathway? What would it take in the way of capital investment? What would

it take in the way of incentives? What would it take in the way of subsidies? What would it take in the way of regulation to move it both in the United States and internationally?

**The European Union has said that it would set a basic target of limiting the effects of climate change to 2 °C.**

**Julian Braithwaite:** It is the best available target at the moment based on the science. I think we then see that that links back to something you can measure, which is about 450 to 550 parts [of carbon dioxide] per million in the atmosphere, which then allows you to start setting things that you can actually target, such as your caps on emissions.

**Michael MacCracken:** It's important, I think, to understand that when Julian says 2 °C, it's 2 °C above pre-industrial levels — not 2 °C above present. And so, as we're almost halfway there and committed to go another quarter of the way, we're very close [to that warming limit] and we need to act very soon. On the CO<sub>2</sub> alone, the effect on the ocean is starting to appear with ocean acidification, and a lot of the oceanographers are getting very worried about what's going to happen to marine life.

**Grumet:** I believe there is actually one scientist among all of our elected members of Congress, so this question of how science engages

with policy is obviously a very important one. I would start with the unfortunate reflection that I think there is a profound disconnect between ecological imperative and near-term political possibility. And the problem that that creates, as the science becomes ever more clear and the impact ever more chilling, is that the debate has shifted among those who are opposed to action from the question of science to now questioning the solution.

**How do we engage the developing world?**

**Pershing:** If you would like to have developing countries engaged, you can't merely tell them 'we're going to make you do it'. You've got to demonstrate that you're going to do it yourself. If we can demonstrate technology potentials, if we can demonstrate the commercial viability of programmes, if we can demonstrate the profitability of these solutions, all of which I think are inherently realistic and plausible, then we can transfer that information.

**Braithwaite:** There are two visions of how we go forward being presented to the large emerging economies in the developing world. One is a voluntary system, in effect based on international peer review. The other is one where you continue to develop the global carbon markets, you continue to develop principles of the Kyoto Protocol where you have mandatory

**"Scientists need to speak out very clearly on the exact details of what the policies are." — Michael MacCracken**



M. TEMCHINE

**"The US needs to take advantage of the learning curve that Europe has experienced." — Jim Rogers**



M. NGAN/AFP/GETTY IMAGES

caps on emissions — you have common but differentiated commitments, which does offer the prospect in the longer term that at some point China and India will have caps on their emissions.

**Grumet:** There is also, I think, a very real question about whether CDM offsets [from the Kyoto Protocol's Clean Development Mechanism, in which developed countries fund projects in the developing world to compensate for their own emissions] will have the environmental integrity that we ultimately need. I think they have a role to play, but I think it has to be a role that is predicated upon a few years of learning.

**Braithwaite:** The European emission-trading scheme and what we've been doing through the CDM is probably the largest real-world example of all of these policies and actions. I think it's fair to say that Europe has been a sort of global laboratory for these policies. And clearly, there are some things that we've learned and we can improve. But I think the point is that in Europe we still think that these systems can provide the developed world with some

**"Outrage among scientists is kind of modest annoyance among the rest of us." — Jason Grumet**

system for offsets in the developing world to bring these nations in, to give them a stake in this global carbon market.

**Rogers:** The European community has done us a real favour by experimenting and

expanding and deploying cap-and-trade, as well as the CDM. And as I have studied some of the issues that they have addressed, these are issues that are not insurmountable. We need to take advantage of the learning curve that Europe has experienced and allow that to give us the courage to build on what it has done going forward.

**We tend to talk about costs, but you're a businessman. There are also opportunities for profit out there.**

**Rogers:** I see this — mainly because our company is so dependent on coal — from a cost standpoint, because 70% of the electricity our consumers use is predicated on coal. But that's on the one hand. On the other hand, I see a great profit opportunity here for technology development. If we can get the business model right for the utility industry, and if we can start pouring billions of dollars into energy efficiency and developing those technologies, I think that offers the greatest hope in the short term as we wait for the development of carbon capture and storage, of battery technology and of other technologies.



**Power struggle:** legislation to curb US emissions will need to strike a political compromise.

#### Should scientists act as advocates on this issue?

**MacCracken:** If you want to avoid dangerous or catastrophic kinds of consequences such as the loss of Greenland, you've got to get on a path where emissions from developed countries are going down by around 80% by 2050. You have to do that. And we'll have to get developing countries to go along as they can, and go down further after that. So I think scientists need to



**Julian Braithwaite:** in an ideal world, scientists shouldn't have to be advocates.

speak out very clearly on the exact details of what the policies are.

**Pershing:** I think that the scientific community has been under-represented in the dialogue and has taken a pass when it should have taken a step forward. It has basically proposed that others know better as to what should be done, and that's not evident. If we take the past 20 years where there has been complete and total inaction, the scientific community in the first IPCC [Intergovernmental Panel on Climate Change] assessment report laid out explicitly the nature of the problem and made proposals as to what ought to be done. Twenty years later, very little has happened. So I suggest the scientific community needs to be much more aggressive.

**Braithwaite:** I think in an ideal world, they shouldn't have to be advocates; their voice should be heard anyway. When their voice isn't being heard, then that's a different situation. I'm not going to comment on the United States, but in the United Kingdom, I think if we tried to put together public policy without basing it on the best available science, we'd get ourselves into trouble very quickly.

**Grumet:** The one other point that I will make is that, in our system, there is such a profound notion of there being two sides to every issue. I think where the scientific community finally rose up with some outrage — and outrage among scientists is kind of modest annoyance among the rest of us — was when the real scientific community was fully convinced of the basics of the ecological reality, but there were one or two folks out there pushing a different [sceptical] side. Yet the situation would be consistently set up as one scientist thinks this and the other scientist thinks something different. And finally I think about a year ago the scientific communities kind of got fed up with that.

**That sounds like it was a problem with the media.**

**Grumet:** Well, of course it's a problem with the media, but sitting here in the National Press Club, I would not be so bold as to suggest that.

**Pershing:** It's also a problem with the scientific community, which has been reluctant to ever come out on any side of any issue. That's not the standard scientific process. There's always room for doubt and uncertainty. But in this particular instance, my sense is the scientific community has done itself a disservice. ■

**Nature reporter Jeff Tollefson moderated this discussion in Washington DC. For a full transcript, visit <http://www.nature.com/news/specials/climatepolitics/index.html>. See also Editorial, page 319.**

S. WALSH/AP

M. TEMCHINE



# What every president should know

If you want to lead the free world, you'd better know your physics. That's the lesson from a popular undergraduate class, called 'Physics for future presidents', taught by **Richard A. Muller** at the University of California, Berkeley. Here he sets some typical questions. An interactive version of this quiz with extended answers is online at [www.nature.com/news/specials/climatepolitics/index.html](http://www.nature.com/news/specials/climatepolitics/index.html).

**1) Electricity from the wall plug costs about 10 cents per kilowatt-hour (kWh). If you were to get the same electricity by buying AAA alkaline batteries at the local store, the cost of that electricity would be:**

- (a) 15 cents per kWh
- (b) 94 cents per kWh
- (c) \$2.50 per kWh
- (d) \$1,000 per kWh

**2) A gram of which of these is most toxic?**

- (a) Botulinum toxin
- (b) Arsenic
- (c) Anthrax spores
- (d) Plutonium dust (inhaled)

**3) The highest achieved efficiency (solar energy converted to electrical energy) of solar cells is approximately:**

- (a) 4%
- (b) 15%
- (c) 28%
- (d) 41%

**4) A typical high-resolution spy satellite has how long to photograph a location?**

- (a) 10 seconds
- (b) 1 minute
- (c) 12 minutes
- (d) 90 minutes

**5) The dose for radiation illness (50% chance of death within a month) is 300 rem, whole body. The dose to trigger on average one cancer is:**

- (a) 2.5 rem
- (b) 25 rem
- (c) 250 rem
- (d) 2,500 rem

**6) Compared with a gallon of gasoline, the energy supplied by a gallon of liquid hydrogen is approximately:**

- (a)  $\frac{1}{3}$  (that is, it has less energy per gallon)
- (b) The same energy per gallon
- (c) 3 times more energy per gallon
- (d) 12 times more energy per gallon

**7) Compared with the energy released when a pound of gasoline is burnt, the energy released when a pound of TNT is exploded is about:**

- (a) 2 times greater
- (b) 13 times greater
- (c) the same, within 40%
- (d) less by a factor of 15

**8) Of the deaths caused by the Hiroshima atomic bomb, the fraction attributed to cancer was:**

- (a) Less than 2%
- (b) About 7%
- (c) About 20%
- (d) More than 50%



**9) A critical mass of plutonium has a volume of:**

- (a) 3 tablespoons
- (b) 1 soft-drink can
- (c) 1 gallon
- (d) 3 gallons

**10) In one computer cycle (a billionth of a second for a slow laptop), light travels about:**

- (a) 1 foot (30 centimetres)
- (b) 300 metres
- (c) 3 kilometres
- (d) 300 kilometres

**11) In the past 100 years, the carbon dioxide level in Earth's atmosphere has increased by what fraction of its previous value?**

- (a) Less than 1%
- (b) 3%
- (c) 30%
- (d) 112%

**12) The rocket that won the X Prize in 2004 achieved an altitude of 100 kilometres. To go into orbit would require more energy. How much more?**

- (a) 1.414 times more
- (b) 2 times more
- (c) 7 times more
- (d) 32 times more

**13) The International Atomic Energy Agency's 2006 estimate for the number of excess cancer deaths expected worldwide from the Chernobyl nuclear accident was:**

- (a) Less than 1,000
- (b) 4,000
- (c) 24,000
- (d) 1.3 million

**14) The ozone layer in the atmosphere is created by:**

- (a) Carbon dioxide
- (b) Sunlight
- (c) Sulphur from fossil fuels
- (d) Chlorofluorocarbon compounds (such as Freon)

**15) Light in a fibre carries more information per second than electricity in a wire because:**

- (a) It has a higher frequency
- (b) It travels faster than electricity
- (c) It makes use of quantum effects
- (d) It doesn't. Wires transmit higher bit rates. (That's why they are used in computers.)

**16) The power in a square kilometre of sunlight is:**

- (a) 1 kilowatt
- (b) 1 megawatt
- (c) 10 megawatts
- (d) 1 gigawatt

**17) To be legal for consumption in the United States, the radioactivity of one litre of ethanol (drinking alcohol) must be:**

- (a) Less than 12 decays per minute
- (b) Below the threshold of standard Geiger counters
- (c) Not measurable by accelerator mass spectrometry (the most sensitive detection method)
- (d) More than 4,000 decays per minute

Richard A. Muller's book *Physics for Future Presidents* will be published by Norton in 2008. Answers on page 346.



## Independent evidence backs call for a badger cull

SIR — In your Editorial ‘In for the cull’ (*Nature* 450, 1; 2007), you implied that my assessment of the scientific evidence relating to bovine tuberculosis (TB) in cattle and badgers was influenced by political considerations. This attack on my integrity as the government’s chief scientific adviser has no foundation whatsoever. The editorial staff of a journal surely understand fully the importance of scientific challenge and debate, as distinct from personal attacks. Yet your Editorial does not address the science.

As you point out, the scientific content of the report on cattle from the Independent Scientific Group (ISG), under the chairmanship of John Bourne, has been published in the scientific literature and hence peer-reviewed, but the conclusions reached by Bourne and his colleagues in the report have not.

My role is to provide independent scientific challenge and advice on important issues, such as this, in which the science is critical. To do my assessment, I assembled a team of independent, well-respected experts who brought international expertise in the necessary disciplines, particularly badger ecology, epidemiology, immunology and bovine TB. I had no idea what conclusions they would reach.

Although the scientific conclusions produced by my experts differ from the main conclusion of the ISG report, they nonetheless follow directly from the ISG’s data, which clearly show that carrying out badger removal over a large area and a sustained period of time, together with cattle removal and other controls, would deliver an overall reduction in TB incidence in cattle herds. This is the only effective course of action until efficacious vaccines become available.

Bovine TB in cattle is the most serious endemic animal disease in the United Kingdom: the ISG reports a doubling of herd breakdown every 4.5 years in the high-incidence area. These data suggest that, in the randomized badger-culling trial, badgers could account for up to 40% of new confirmed incidents of bovine TB in cattle.

On the basis of the scientific evidence, I do not believe that we can control TB in cattle — and badgers — without removing the sources of infection in both species. Other countries have been unable to control TB in cattle without addressing the wildlife reservoir (N. E. Tweddle and P. Livingstone *Vet. Microbiol.* 40, 23–39; 1994).

I utterly reject any suggestion that I was — or could have been — influenced in my science advice by farmers, policy-makers or

politicians. My mantra of openness, honesty and transparency continues unabated.

**David King**

Government Office for Science, Department of Universities, Innovation and Skills, Kingsgate House, 66–74 Victoria Street, London SW1E 6SW, UK

**Nature has no reason to question Sir David’s honesty, nor any evidence of political doctoring of his report. We stand by the concerns about the practice of scientific advice raised in the Editorial — Editor, *Nature***

## Kyoto: talks must include key aspects of science

SIR — Gwyn Prins and Steve Rayner are correct to argue, in their Commentary ‘Time to ditch Kyoto’ (*Nature* 449, 973–975; 2007), that the Kyoto Protocol has failed and needs a radical rethink, but they do not diagnose the roots of this failure correctly. Lessons are less likely to be learned without this diagnosis.

There were two fundamental problems: one inherent in the negotiating model and the other in the issue structure that came to dominate the process. The first overestimated the possibility that science would counter the divergent interests of different states sufficiently for them to act in the greater global good; the second diminished the possibility that force of moral obligation would enhance this prospect.

A better understanding of the science might inform post-Kyoto negotiations more productively. For example, the logarithmic nature of carbon dioxide forcing, with each additional tonne having a smaller effect than the last, suggests that burdens should be less for countries industrializing now than for those that industrialized 100 years ago or more. And the ‘Hansen alternative scenario’ (J. Hansen *et al. Proc. Natl Acad. Sci. USA* 97, 9875–9880; 2000) suggests that much might be done by mitigating carbon-forcing agents, such as carbon soot from inefficient biofuel combustion, because this would bring enormous co-benefits from addressing indoor-air pollution in India and elsewhere.

‘Minilateralism’ in groups such as the G8+5 Climate Change Dialogue and the Asia-Pacific Partnership on Clean Development and Climate (AP6), recently joined by Canada, is likely to be more productive than full multilateralism. But neither approach is going to be particularly productive unless key aspects of climate science inform the negotiations, and the normative discourse makes accurate and credible demands on parties such as the United States and Australia, which are unlikely to set aside their interests in the absence of such measures.

**Aynsley Kellow**

School of Government, University of Tasmania, Private Bag 22 Hobart 7001, Australia

## Kyoto: no time to rearrange deckchairs on the *Titanic*

SIR — Gwyn Prins and Steve Rayner, in their Commentary (*Nature* 449, 973–975; 2007), manage to be perfectly right and utterly wrong at the same time. Their criticism of the bureaucratic Kyoto Protocol is justified on many crucial points (although they don’t mention that the physical impact of the protocol on the climate system would be negligible even if it worked). The novelty of this summary of well-known deficiencies in the treaty is that the list comes from independent European scientists rather than White House mandarins. Is there anything substantially new beyond that provocation?

Yes, in the sense that Prins and Rayner boldly propagate a “bottom-up ‘social learning’” approach to climate policy that aspires to “put public investment in energy R&D on a wartime footing”. I agree with the importance of both elements to twenty-first century climate protection, but doubt whether there is a solid causal chain linking them. Fine-scale measures and movements towards sustainability, as well as technological and institutional innovation strategies, are needed to decarbonize our industrial metabolism and to force policy-makers to face the challenges ahead. Fancy phrases such as “the silver buckshot” may help to sell the case.

Time is crucial, however. It is unlikely that a bottom-up, multi-option approach alone will be able to mobilize war-level climate-protection efforts by all the major emitters (including Russia, China and India) within the one or two decades left to avert an unmanageable planetary crisis. Without a ‘global deal’ — designed for effectiveness, efficiency and fairness and providing a framework to accommodate every nation — there will be neither sufficient pressure nor appropriate orientation towards the climate solutions we desperately need. The bottom-up and top-down approaches are complementary and must be pursued interactively.

Kyoto is simply a miserable precursor of the global regime intended to deliver genuine climate stabilization — and was never expected to be more. “Ditching” it now would render all the agonies involved completely meaningless after the event, denying the entire process of policy evolution the slightest chance to succeed. So, instead of rearranging the deckchairs on the *Titanic* through social learning, let us ditch pusillanimity.

**John Schellnhuber**

Potsdam Institute for Climate Impact Research, PO Box 601203, D-14412 Potsdam, Germany

See *Nature’s* climate blog at <http://tinyurl.com/37g2pk>. Readers are welcome to add their comments.

**Answers to ‘What every president should know’**

1 d; 2 a; 3 d; 4 b; 5 d; 6 a; 7 d; 8 a; 9 b; 10 a; 11 c; 12 d; 13 b; 14 b; 15 a; 16 d; 17 d

## COMMENTARY

# Who has the ear of the president?

50 years after the appointment of the first presidential science adviser, the White House is flooded with scientific information. **Roger Pielke Jr** suggests how the next administration might develop ways to use it best.

**O**n 15 November 1957, as part of his response to the Soviet launch of Sputnik, President Dwight Eisenhower swore in James Killian, president of the Massachusetts Institute of Technology in Cambridge, to the newly created position of special assistant to the president for science and technology. Since then, 14 men — almost all physicists — have served 10 presidents as ‘science adviser’, as the position is more commonly known.

A recent article in *Physics Today* looked back wistfully on the position’s early years<sup>1</sup>: “Never before or since have scientists had a firmer influence on the reins of power that direct national policies.” Recommendations that accompany such nostalgia, perhaps most evident during the term of the current and longest-serving science adviser John Marburger, draw more from legend than from history, with far more attention paid to how science advice is given rather than to how it is used.

The president today receives an almost overwhelming flood of expert advice — from government committees, national academies, national and international scientific organizations, self-organized scientific groups, advocacy groups, and many others. By some estimates, the activities of more than 18,000 governmental science and technology advisers generate 6 new reports every day. The president is thus hardly in need of more scientific advice. The central challenge of managing expertise in today’s government is to make effective use of the flood of information that threatens to drown decision-makers.

## Lessons from the past

During 2005 and 2006, I interviewed seven science advisers who served under presidents from Lyndon Johnson to George W. Bush<sup>2</sup>. Although the role of each science adviser was as different as each president, some characteristics remain constant. Neal Lane, the second science adviser to President Bill Clinton, explains that: “You have to support the president in his policies, whatever those policies are, and you have to do that while preserving your integrity as a scientist and the integrity of your office. If the president says something that’s just wrong scientifically, then you’ve got to figure out how to explain to the American people what the president really meant to say

**“You have to support the president in his policies, whatever those policies are.”**

— Neal Lane



President Eisenhower (right) chats with the inaugural science adviser James Killian at a dinner in 1959.

in such a way that everybody is okay.”

Contrary to popular wisdom, the role of science adviser was in decline from the moment that President Eisenhower tellingly rushed through Killian’s appointment ceremony in order to depart for an eagerly anticipated golf vacation in Augusta, Georgia. And in his 1961 farewell address, Eisenhower expressed considerable wariness about scientific expertise in government: “In holding scientific research and discovery in respect, as we should, we must also be alert to the equal and opposite danger that public policy could itself become the captive of scientific-technological elite.”

Yet as the adviser’s influence has declined, scientific and technological expertise at the highest levels of government has been triumphant. William T. Golden, investment banker, philanthropist and a chief architect of the science-adviser position, wrote in 1950 that the government could draw on “somewhere between 20 and 200” top scientists. By 2003 there were approximately 8,000 scientists serving on about 400 federal advisory committees. Without effective mechanisms to turn advice into options, and options into action, the often

heroic efforts of these scientists will amount to little more than academic exercises.

As science and technology have become more important to governmental decision-making over the past half century, one result has been the reduced importance of a personal science adviser to the president. Because policy issues with scientific or technical content are ubiquitous, it is unrealistic to think that one person with a small staff can hope to serve as the president’s sage.

In any case, a president does not make scientific judgements. A president makes political judgements that may involve scientific or technical considerations in the evaluation of alternatives for action. Jack Gibbons, President Clinton’s first science adviser, explains: “Science is not an overarching national goal for the president. It’s only as it serves to help achieve these larger goals that science takes its place in the crown of important activities for the president.” Science is a determining factor in decision-making only when there is very little political conflict. By contrast, intense political conflict magnifies small differences over science as a proxy for political debate, making the management of expertise all the more important to prevent experts simply cancelling each other out<sup>3</sup>.



Jack Gibbons speaks at a press conference in 1992. He later became science adviser to President Clinton.

Structural factors further limit the role of the science adviser in presidential decision-making. In 1976, Congress reversed President Richard Nixon's earlier decision to abolish White House science advice when it formalized the Office of Science and Technology Policy (OSTP) in legislation and with it created a second role for the president's science adviser as OSTP director. Gibbons says that this legislation, perversely, meant that the science adviser would never be among the 'inner circle' of White House advisers because, unlike other special assistants to the president, the OSTP director could be asked to give Congressional testimony. Calls for institutional reform to bring the science adviser closer to the centre of White House decision-making, such as by reinstating the President's Science Advisory Committee<sup>4</sup>, are therefore unlikely to succeed.

### The inner circle

Personal relationships are also an important factor. Lane emphasizes that key political decisions were often made by the president's closest staff, most of whom forged their relationship with the president in the heat of election campaigns. It is telling that Jerome Wiesner, who served under President John F. Kennedy, seems to be the only science adviser who participated in a presidential candidate's successful election campaign. A fundamental political reality is that decision-making at the highest levels of the White House will always be based on the idiosyncrasies of the president and his or her staff regardless of the method by which science advice is delivered.

Yet the science adviser is neither irrelevant nor powerless. The OSTP works with the Office of Management and Budget, which prepares the president's budget for the entire government. "In today's OSTP, we set our work schedule and products deliberately to synchronize with the budget cycle," says Marburger. Coordination of research and development

budgets across federal agencies is therefore central to the job. No other cross-department area of government activity has such a unique input into the budget process, which may help to explain why research and development spending has increased steadily for decades. The downside of this unique role is that the science adviser and the OSTP must be careful to avoid having their views discounted as coming from an internal 'science lobby'.

In balancing these challenges, science advisers have adopted vastly different roles. George Keyworth, who served under Ronald Reagan, says he found his niche by focusing his advice solely on the Strategic Defense Initiative. Ed David recounts how he was enlisted to help President Nixon's re-election chances by having NASA reschedule its Apollo moon launches away from the election. Donald Hornig, President Johnson's adviser, related that science was often instrumental in international relations as "a wonderful lubricant for foreign policy initiatives". Such differentiated responsibilities and experiences challenge the notion of the science adviser as the president's sage. Whoever is next appointed to the advisory post will no doubt shape the relationship with the president in his or her own unique way.

Even if the science adviser has rarely, if ever, held the 'reigns of power' over the past 50 years, scientific and technical expertise has nonetheless triumphed in government. Historian Daniel Kevles at Yale University in New Haven, Connecticut, calls the contemporary political environment "unbelievably pluralistic" in which "there is hardly an issue you can think of that doesn't turn to some extent on technical knowledge". Thus, the challenge of enabling effective science advice under the next president will not be met by identifying one person or even a committee to serve as a fountain of

wisdom on all matters scientific or technical. Such quaint notions ignore the lessons and realities of science advice in government.

The reality of pluralistic policy-making helps to explain why today so many issues involving science are politicized, and will continue to be so, under all future presidents. The scientific community can assist the next president by focusing greater attention on the overwhelming supply of expert advice beyond the White House that feeds into all aspects of government decision-making<sup>5</sup>. In practical terms, this would mean eschewing calls to separate science from politics, and fostering instead more sophisticated ways to integrate science with the needs of policy-makers.

### The triumph of expertise

Scientists in government need more effective means to elicit from decision-makers the policy-relevant questions that need to be addressed by scientific and technical experts, building on the experiences of the congressional Office of Technology Assessment.

Another useful strategy would be to ask some expert advisory committees to go beyond the discussion of technical matters by presenting a wide range of policy options to decision-makers. This would require thinking about scientific advice and its implications more comprehensively, because discussion of policy options requires integrative, interdisciplinary knowledge. The science adviser (and the OSTP) might also contribute to this process by serving as an in-house 'think tank' for the president, a function the social scientist

Daniel Yankelovich calls an "options czar". Presenting options would help preserve the public credibility of the science adviser by clearly delineating the differences between advice, advocacy and decision-making<sup>6</sup>.

Hopes within the science community that the next president will somehow return the science adviser to a position of power are based on unrealistic expectations. The relationship between the next president and his or her science adviser will be as unique and idiosyncratic as those under the past ten presidents. Far more important for effective decision-making will be how the next administration manages and uses the vast infrastructure of expert advice that it will inherit.

■ Roger Pielke Jr is at the Center for Science and Technology Policy Research, Cooperative Institute for Research in Environmental Science, University of Colorado, Boulder, Colorado 80309, USA. e-mail: [pielke@colorado.edu](mailto:pielke@colorado.edu)

**"A president does not make scientific judgements. A president makes political judgements that may involve scientific or technical considerations."**

1. Rigden, J. *Phys. Today* **60**, 47–53 (2007).
2. <http://sciencepolicy.colorado.edu/scienceadvisers/>
3. Sarewitz, D. *Environ. Sci. Pol.* **7**, 385–403 (2004).
4. Garwin, R. *Nature* **449**, 543 (2007).
5. Jasanoff, S. *The Fifth Branch: Science Advisors as Policymakers* (Harvard Univ. Press, Cambridge, 1990).
6. Pielke Jr, R. *The Honest Broker: Making Sense of Science in Policy and Politics* (Cambridge Univ. Press, 2007).



## BOOKS &amp; ARTS

## A sixth mass extinction?

Past species losses have much to teach us about current and future declines due to human activity.

**Terra: Our 100-Million-Year-Old Ecosystem — and the Threats that Now Put it at Risk**

by Michael Novacek

Farrar, Straus and Giroux: 2007. 480 pp. \$27

### Chris D. Thomas

This year the baiji river dolphin (*Lipotes vexillifer*), a victim of the pollution and boat traffic of China's Yangtze river, was added to the list of creatures on the verge of extinction. Is this part of the sixth mass extinction in 450 million years, or does the recent spate of losses caused by humans represent a blip in the history of life on Earth? Michael Novacek's *Terra* takes stock of the situation and provides an opportunity to learn from the past.

Novacek ambitiously amalgamates the comings and goings of species over the past 400 million years to set the stage for his description of recent human-induced extinctions and an examination of today's threats to species. I would have preferred a single, persuasive synthesis rather than a hotchpotch of stories relating to the past, present and possible future, but the author covers his material engagingly, interspersing it with personal anecdotes.

I do not buy the thesis that the proliferation of flowering plants and associated insects about 100 million years ago was more important in determining today's life and ecosystems than were other events in the geological past. After all, land covers only 29% of Earth and flowering plants are not the only terrestrial life forms. If their evolution is so important, are flowering plants more or less sensitive to human onslaught than other components of global ecosystems, and how is their geological history relevant to the answer?

It would have been useful to include critical comparisons of past, recent and future extinction rates. How do the numbers stack up? Three of the recognized 'big five' mass-extinction events occurred in the past 300 million years, each accounted for the extinction of some 30% or more of marine genera. If 30–50% of present-day species are on their way to extinction, to use the numbers that Novacek quotes, then roughly 15–25% of genera might be lost — still short of the big-five estimate.

There are many difficulties in making comparisons between geological data and future projections, the assumptions and uncertainties involved merit a serious discussion missing here. Projected extinctions are, on the face of it,



The baiji river dolphin, *Lipotes vexillifer*, of China's Yangtze river may soon disappear.

already within the range of the largest extinction events of the past 300 million years, each of which has heralded the arrival of a major new geological period or epoch. The human-dominated 'anthropocene period' seems to be unfolding.

Limiting the rate of extinction will be extremely difficult because space and resources are finite. Take the demand for food and bio-energy production: about a quarter of maize (corn) production in the United States for 2007 is designated for biofuel. To fulfil the new US 2017 target for biofuels from this inefficient source would require about 0.5% of Earth's land surface, and a considerably greater percentage of its potential farmland. If the same area of uncultivated land is taken into cultivation elsewhere to replace lost food production, this will drive about twice as many species to extinction — through extra habitat loss — as would be reprieved by mitigating climate change. It has been shown that if just 6% of the land currently used for maize-ethanol production for biofuels were to be displaced into further tropical deforestation — releasing the carbon held in those forests — no climate-change benefits from carbon mitigation would accrue at all.

Win-win solutions are sometimes possible,

for example, by saving carbon in intact forests that protect biodiversity. But we cannot increase the world population by 50%, and increase the food resources available to every citizen, and grow our energy, pharmaceuticals, plastic and other products, and still maintain existing ecosystem services and all extant species. Of course, we shall solve some of these issues with technological fixes. Yet if we maintain 9 billion avaricious people on Earth for the next millennium, a sixth extinction event seems inevitable.

The geological perspective of *Terra* is bizarrely reassuring. Humans will presumably be gone within a few million years, perhaps sooner. If the past that Novacek describes is a guide to the future, global ecosystem processes will be restored some tens of thousands to a million years after our demise, and new forms of life over the ensuing millions of years will exploit the denuded planet we leave behind. Thirty million years on, things will be back to normal, albeit a very different 'normal' from before. It is good to be optimistic. The problem is living here in the meantime. ■

Chris D. Thomas is professor in the Department of Biology, PO Box 373, University of York, York YO10 5YW, UK.

# Prescient marine champion

## **The Human, the Orchid and the Octopus: Exploring and Conserving our Natural World**

by Jacques Cousteau & Susan Schiefelbein  
Bloomsbury USA: 2007. 320 pp. \$25.95

### **Ken Collins**

Jacques Cousteau turned the television screen into a porthole onto the underwater world and each of his documentaries attracted some 250 million viewers worldwide. It was his idea to put a yellow stripe down the arms and legs of early wetsuits to make divers more visible on film. He was the reason I became a marine scientist.

Ten years before his death in 1997, Cousteau wrote *The Human, The Orchid and The Octopus*, now available for the first time in English, in collaboration with Susan Schiefelbein. She had worked alongside him and written the scripts for many of his films, and has since added a biographical introduction that gives us an insight into the development of Cousteau's views on conserving the marine environment, which forms the main subject of the book. It is a very readable mix of stories of underwater exploration, overlaid with an increasing awareness of environmental issues and an emerging philosophy that will be widely appreciated.

The book's mysterious title derives from the last chapter — 'The miracle of life' — which discusses the most highly developed examples of vertebrates, plants and invertebrates. In this chapter, Cousteau's fascination with the complexity of marine life is illustrated by his delight in discovering and understanding the behaviour of small fishes — for example, the habit of the perch-like wrasse of laying its eggs in the bowls of sponges. He was tickled by an example of native wisdom in Indochina during his early navy career, when a rower slid over the side of their boat to grab edible evidence that the local fish took a midday siesta. To satisfy his huge curiosity, Cousteau undertook many hair-raising dives, including an early, extremely deep journey in a bathyscape submersible; his definition of a successful dive soon came to be one that returns to the surface.

Wonder at the marine environment led to his concern for its protection, long before it became a global issue. He publicly renounced spear-fishing in 1960, proclaiming that "we must have a fundamental respect for nature". Soon afterwards, he mobilized 11,000 compatriots to prevent a train from delivering its load of radioactive waste for dumping in the Mediterranean Sea. Later, the Madrid Protocol was drawn up as a direct result of his personal plea to world leaders for international protection of the Antarctic continent.

As an experienced marine scientist and director of Monaco's Oceanographic Museum, Cousteau fully understood the risks posed by



Jacques Cousteau, centre, was a pioneer in protecting the seas from human harm.

pollutants. In the book, this awareness is presented alongside discussion of the world's scriptures, in which water is revered for its bounty and its purity. As Cousteau became more caught up in conservation issues, he realized that policy decisions are too frequently dictated by market, rather than human, values. Schiefelbein adds an epilogue updating many of the environmental issues discussed.

Despite the age of the core text of *The Human, The Orchid and The Octopus*, the issues discussed are still topical. Cousteau observes that shallow coastal waters are the most fertile, but also the most vulnerable to human activities. Problems with world fisheries continue to increase. Even at the time of writing, he noted

that fish could never feed the world.

Cousteau decries the militarization of science through the cold-war years, which produced the newspaper headline "Fear of Peace Depresses Market". He concludes with the thought that humanity's enemy has never been science, but rather that we need to use science to civilize civilization.

In his foreword, Bill McKibben considers fear to be the second-best motivator for conserving our planet — the most powerful being love for it, as generations learned from Jacques Cousteau.

Ken Collins is a senior research fellow at the National Oceanography Centre, Southampton SO14 3ZH, UK.

## Human distilleries

### **Rethinking Expertise**

by Harry Collins & Robert Evans

University of Chicago Press: 2007. 160 pp.  
\$37.50

### **Robert P. Crease**

The modern world depends on experts and expertise. We use them all the time, without a second thought, in our personal and professional lives. Yet controversies over such matters as the causes of AIDS and global warming, and a growing public distrust of science, provoke scepticism about the abilities and agendas of so-called experts. According to Harry Collins and Robert Evans, the notion of expertise

needs an overhaul.

The authors of *Rethinking Expertise* are both sociologists. Collins came across the topic during his 30-year study of scientists involved with gravitational waves, which culminated in his book *Gravity's Shadow: The Search for Gravitational Waves*. He decided that the portrayal of experts as knowledge producers who pass on nuggets of truth to clients was woefully inadequate, a point he and Evans presented in a 2002 article reprinted in *The Philosophy of Expertise* (an anthology that I co-edited). *Rethinking Expertise* expands on that work, calling for "a new sociology of expertise".

Collins and Evans argue that expertise does



not involve passing on truths from producer to consumer, but takes the form of fallible judgements made in situations in which all the desired information is not available. This applies to scientific collaborations and governmental projects. "The speed of politics," they write in a characteristically succinct and clever remark, "exceeds the speed of scientific consensus formation." This leads them to defend their seemingly pedestrian conclusion that governments should rely on experts as "the best way to distil human experience of an uncertain world" — even when those experts have no final answers.

The authors needed to defend themselves. Some sociologists sharply attacked their 2002 article on the grounds that the technical and political phases of a controversy cannot be neatly separated, and that governments often use expert advice on technical questions ("Does activity X meet our standards for safety?") to disguise political questions ("Should we do activity X?"). This problem is best met, critics contended, by expanding public involvement in technological decision-making, breaking down the boundaries between experts and laypeople.

*Rethinking Expertise* argues the reverse, namely that disguised political questions would be combated more effectively by a more

rigorous understanding of whose technical expertise really should weigh in on technological decision-making. This is what Collins and Evans call the "problem of extension", or "who is entitled to contribute to the technical component of a technological decision". If an extension is too technocratic, it will foster distrust; if it is too broad, there is a risk of invoking 'technological populism', which means decisions could be overly influenced by politics or lifestyle choice.

The authors address the problem of extension by offering a "periodic table of expertise" — an attempt "to classify all the types of expertise that might be brought to bear on a technological problem". These include about a dozen varieties, ranging from language-speaking skills to higher, domain-restricted categories. There is 'contributory' expertise from active practitioners of a field (the usual concept of an expert) and 'interactional' expertise, which involves fluency in a field in the absence of contribution. Sociologists and journalists have interactional expertise, and it is widespread in science (in which scientists must interact across disciplines), in management and in peer-reviewing.

Interactional expertise is one of the authors' most novel ideas. They dramatize it by reporting the results of several Turing-test-like imitation games. In one, judges could not distinguish

between colour perceivers and colour-blind participants, probably because the latter had been immersed since birth in the language of colour and acquired the ability to pass as perceivers. In another, judges could not pick out Collins in a group of bona fide gravitational-wave researchers.

These results have startling implications. For instance, Evans has shown that minorities can acquire enough interactional expertise of mainstream culture to foil attempts to pick them out by questioning alone. This suggests that if minorities live by different values, it is not through ignorance but choice.

Collins and Evans put their points vividly, with elegant language and diagrams. They admit that there is more to technological decision-making than "sorting out the appropriate groups of experts" and that they are only addressing the technical phase of public controversies. They modestly claim to be only setting "the ball rolling".

Their book starts to lay the groundwork for solving a critical problem — how to restore the force of technical scientific information in public controversies, without importing disguised political agendas.

Robert P. Crease is a professor in and chairman of the Department of Philosophy, Stony Brook University, New York 11794, USA.

## Blurring our edges

### Enhancing Evolution: The Ethical Case for Making Better People

by John Harris

Princeton University Press: 2007. 260 pp. \$27.95, £16.95

### Ending Aging: The Rejuvenation Breakthroughs That Could Reverse Human Aging in Our Lifetime

by Aubrey de Grey and Michael Rae

St. Martin's Press: 2007. 400 pp. \$26.95

### Judy Illes

Two new books explore the potential for improving the human condition, one from a medical-ethics viewpoint and the other from a more hands-on biogerontological perspective. Each identifies physical health and cognition as the key areas for attention in the short term, and each speculates on future possibilities for life extension — going well beyond what might be imaginable, or ethical, today.

John Harris's book *Enhancing Evolution* discusses the future of humankind and of research, and looks at issues such as reproductive choice, designer children and disability. Harris systematically analyses historical and contemporary writings to support his central claim that the pursuit of methods to enhance the human condition is a moral obligation. His treatment of opinions by conservative thinkers and phi-



Could we all live longer than Japan's Yone Minagawa, who reached her 114th birthday?

KYODO NEWS/AP PHOTO

losophers such as Leon Kass, Jürgen Habermas and Francis Fukuyama is at times abrasive. But, whether or not one agrees with Harris's position, he dissects their opposing claims with impressive effectiveness. He makes a persuasive case that today's biotechnologies — including neuroengineering, stem-cell research and cloning to improve bodies and brains — are on the continuum of an age-long pursuit by humans to improve themselves and therefore, he argues, are permissible and morally essential.

I question the author's contention that choices such as non-medical sex selection can be as straightforward as picking a white shirt over a Hawaiian one, or that "inevitability is the only reason dying of old age is acceptable". But I do agree that everyone — scientists, ethicists and the public — has a role in the research enterprise, not least because of benefit sharing. That is not to say that we should all be drafted into research, and certainly no end would justify exploitation of children and people with diminished abilities, but service to society by its members is an honourable goal.

Aubrey de Grey and Michael Rae's *Ending Aging* is a more 'new-wave' treatment of enhancement, longevity and immortality. Through chapters with catchy titles, such as 'Putting Zombies to Rest', and gripping language — "I shall look forward to shaking your hand in a future where engineered senescence is a reality" — the authors call for research that will deliver a life expectancy to match that of Methuselah, the oldest of the people whose age (969 years) is mentioned in *The Bible*.

*Ending Aging* guides the reader through a maze of advances in molecular and cellular biology that could lead to anti-ageing therapies, which the authors term SENS (for 'strategies for negligible senescence'). The targets are a range of cancer-causing nuclear and mitochondrial mutations, intracellular and extracellular debris, and molecular crosslinks that contribute to pathological decay over a lifetime.

The authors rather unnecessarily brand ageing as repugnant and a curse, and use their book to preach on fund-raising opportunities. By contrast, de Grey's 2005 publication on 'Life extension, human rights, and the rational refinement of repugnance' in the *Journal of Medical Ethics* (co-edited by Harris) is still an edgy but more dignified treatise.

The freedom to pursue ways to enhance human mental and physical capacities and to eliminate negative aspects of the human condition, such as suffering and death, is a fundamental tenet of the trans-humanist movement. Although seemingly worthy, there are problems ahead for the futurists, including for Harris, de Grey and Rae. At what age should life-prolonging measures begin — *in utero*, when development of the nervous system peaks in the mid-20s, or much later? What would become of natural rhythms in wisdom and folk psychology, and how will the roles of elders be redefined? Would we immortalize some and punish others by withholding

life-prolonging intervention — reconfiguring the term 'life sentence' and introducing new systems of justice for the parallel mortal and immortal populations?

Will people be able to step off the immortality train when they are ready? Will suicide increase if we move life expectancy into the hundreds? We cannot remove longevity like a pair of spectacles. And abuse is reported by 1–10% of old people surveyed, mostly directed against women. If society shows increasingly scant regard for centenarians, how will it treat people of 1,000 years old?

Let's not throw away today for tomorrow. *Ending Aging* is likely to appeal to those already converted to the authors' views, and

perhaps will find some traction among those who are more curious than interested in deeper scientific engagement. I end unconvinced by Harris's view, but my interest was piqued by the diverse ways in which physical and mental health, autonomy, culture, and respectful and caring community systems can be realized. After we have maximized human well-being and compressed morbidity to minimize suffering (see A. Kalache *et al. Bull. World Health Organ.* **80**, 243–244; 2002), then let's think about life extension, step-by-step, one decade at a time. ■

Judy Illes is professor of neurology at the National Core of Neuroethics, University of British Columbia, Vancouver V6T 2B5, Canada.

## Archimedes' secrets revealed

### The Archimedes Codex

by Reviel Netz & William Noel

Weidenfeld & Nicolson/Phoenix/  
Da Capo Press: 2007. 336/336/320 pp.  
£18.99 (hbk)/£8.99/\$27.50 (pbk)

### Brian Clegg

In April 2004, biologist Lewis Wolpert won a debate at the Royal Institution in London entitled 'Who was the first scientist?' He championed Archimedes. Although I reckon Roger Bacon comes closer to deserving that label, the result reflects the respect still felt for Archimedes' work. Yet our knowledge of his output has never been complete. *The Archimedes Codex* describes the recent uncovering of text from the Archimedes canon not read since the parchment forming the codex was reused to make a prayer book in 1229.

Only three of Archimedes' books in the original Greek are known, two of which have been lost since first being transcribed. The third, this

codex, had never been fully deciphered. Access to the earliest known version (from the tenth century) was important because of a recent shift in interpretation of Greek texts. The Greeks approached mathematics visually. The diagrams and the exact wording used are crucial in understanding the thinking of a Greek mathematician — yet when the lost Archimedes manuscripts were transcribed, no thought was given to capturing the visual nature of the information.

In *The Archimedes Codex*, Reviel Netz and William Noel contribute alternate chapters, with Noel describing the history of the codex and the work undertaken to decipher it, and Netz exploring Archimedes' mathematics. This discontinuous style is initially unnerving — both authors write in the first person, and it isn't clear that the 'I' of chapter 1 is not the same person as the 'I' in chapter 2. But it soon settles into an engaging account.

Noel, director of the deciphering project, explains the intricacies of dealing with a manu-



Analysing the Archimedes palimpsest with different wavelengths of light.



## EXHIBITION

# The liquid of life

Josie Glausiusz

Atlanta, Georgia, will run out of drinking water within the next four months if it doesn't rain soon. The city is experiencing its driest year since 1931. Only a dramatic cut in water use can save the situation, officials announced in October. The average resident, like the average North American, uses between 227 and 340 litres of water per day.

Most of us would quickly use less water if we had to carry that much for miles every day on our heads in the type of brass jug toted by women in Jalalabad, Afghanistan. One such flower-etched vessel is on display at the American Museum of Natural History's new exhibition, *Water: H<sub>2</sub>O=Life*, and visitors are invited to lift it. Full, the jug weighs about 11 kilograms. I could barely raise it above my knees, let alone hoist it onto my head. Trudging home with hefty pitchers like this is a way of life for many: in some parts of sub-Saharan Africa, women spend between 15 and 17 hours a week collecting water.

*Water: H<sub>2</sub>O=Life* clearly conveys how vital water is to our existence, how imperilled we are without it, and how big the divide is between the affluent West, where water flows freely from taps at a cost of about a quarter of a US cent per litre, and poorer regions of the world, where shanty-town dwellers pay a premium for trucked-in water. State-of-the-art interactive video displays, the museum's dearly loved dioramas (one of a life-size polar bear on ice), maps and satellite images detail grim statistics. For example, 20 litres per day is the minimum amount of clean water required to meet basic human needs — drinking, cooking and hygiene. Some people in parched places have to survive on just 5 litres or less. One quarter of Mexico City's residents have no access to tap water; others have only 1 hour a week.

The exhibit's emphasis is on water conservation, often at the expense of the esoteric science of water or the politics of its



Bottle perspective: 3 litres of source water are used to make 1 litre of bottled water.

scarcity in many conflict zones. Information is largely pitched at the curious layperson or schoolchild — the show was instigated when a 2006 survey by the National Oceanic and Atmospheric Administration revealed alarming public ignorance. The beautifully designed displays excel at showing visitors how to save water and energy. A mound of plastic bottles explains that it takes 3 litres of water to produce a 1-litre bottle of the stuff; and that 40% of them in the United States are filled with filtered tap water — at 1,000 times the price.

The exhibit also artfully demonstrates the innovative methods that people are using to extract, conserve and purify water. More than 700 PlayPump water systems have been installed in rural South Africa and other sub-Saharan countries. Children spinning on these colourful merry-go-rounds 16 times per minute generate enough energy to pump about 1,400 litres of water per hour from boreholes 40 metres deep. (I enjoyed hand-whirling a miniature PlayPump to fill a toy jerrycan from a tiny tap.) Perth in Australia gets 20% of its water supply from the Indian Ocean, desalinated with wind power. Windhoek in Namibia reclaims up to 30% of its water from sewage.

Some water-purification methods on display

are delightfully low-tech. Bangladeshi women have cut cholera incidence in half using sari fabric, folded eight times, to filter bacteria from water. This, more than a century-and-a-half since the link was first made between cholera and polluted water. That connection prompted UK authorities to build separate water mains and sewerage tunnels emptying downstream into the Thames river, a system that is still in use. The Thames and Lee rivers now supply drinking water to 8.3 million residents in the London area every day. Alas, modernity does not always equal progress. Many times a year the antiquated system spews hundreds of millions of litres of untreated sewage directly back into the rivers. With some stretches of pipe more than 100 years old, the city itself accounts for 25% of water leakage in Britain. Whether in sodden London or arid Atlanta, saving water is a hard habit to learn, it seems.

Josie Glausiusz is a New York-based science journalist and editor, and author of *Buzz: The Intimate Bond Between Humans and Insects*.

***Water: H<sub>2</sub>O=Life* runs at the American Museum of Natural History in New York until 26 May 2008. It will then tour the United States, Australia and Canada until 2011.**

script of this age and of uncovering the hidden text beneath the overwriting. His contribution has a personal tone ideal for the palimpsest story but rather too lightweight for scientific technicalities such as the mechanism of the multispectral imaging used to bring out the obscured text. Netz, a classicist at Stanford University, makes less allowance for his audience. His valuable examples may warrant several re-readings for non-mathematicians.

The most exciting of the discoveries in the newly deciphered text was Archimedes' use of infinity. The Greeks were wary of infinity, a concept they endowed with implications of chaos and disorder. When infinity was con-

sidered in Greek mathematics, it was treated as potential infinity, rather than the real thing. A potentially infinite pile of logs will never run out — there are always more — but it contains a finite, if indefinite, number.

Before the codex was deciphered, it was thought that, apart from some playful consideration of true infinity by Galileo Galilei, the concept was hardly touched on until the nineteenth century. Netz painstakingly retrieves text from the codex proposing that two infinite sets have the same size because the elements in them can be put in a one-to-one correspondence. Such sets are now said to have the same 'cardinality', the modern concept that estab-

lishes that two sets are equivalent in magnitude. Yet here was Archimedes using this argument more than 2,000 years before Georg Cantor added it into the mathematical armoury.

Netz also shows how Archimedes used a remarkable physical extension of geometry to calculate the area under a parabola, and undertakes some work based on a game called the *stomachion* (bellyache). However, it is Archimedes' use of the infinite that has the biggest impact on our understanding of the history of mathematics and that best demonstrates the value of Netz and Noel's work. Brian Clegg is the author of *A Brief History of Infinity* and editor of [www.popularscience.co.uk](http://www.popularscience.co.uk).

# Use the calm between the storms

To save lives and livelihoods, natural and social scientists must work with decision-makers and politicians in the time between natural disasters as well as during them.

**Steve Sparks**

Science has many roles in the forecasting and mitigation of natural hazards. It underlies technological solutions to giving early warning, producing advice for authorities in areas at risk, and designing effective relief strategies. It also provides critical information for policy-makers and the public to help save lives and avoid economic losses. The fulfilment of these roles is in practice complex, and has not been entirely successful — as illustrated by recent events such as the 2004 Asian tsunami, the 2005 Pakistan earthquake, and Hurricane Katrina, which hit the United States in 2005. The time has come to do better as environmental change and population growth will increase vulnerability with the potential for many more disasters.

Most human and financial resources are expended in response to crises rather than in their prevention, and much energy is spent on shifting blame. Political interest in natural hazards is at its highest during and shortly after a disaster — when it is too late. Funding for prevention measures and preparedness is hard to come by when there has not been an arresting earthquake or volcanic eruption for a while, even though prevention is much more cost-effective.

In a crisis, policy-makers are under pressure to be seen to be doing something. Poorly conceived or expensive initiatives are announced and enacted, sometimes with unintended consequences. Then initiatives are quietly forgotten as the crisis declines and political agendas move on.

Take the Asian tsunami. Immediately afterwards, many initiatives were proposed, such as the installation of a technological early-warning system for the Indian Ocean, similar to the existing one in the Pacific. Although this system is now progressing through the auspices of the Intergovernmental Oceanographic Commission, many coastal communities in southern Indonesia remain vulnerable to earthquakes and tsunamis. They are too close to the active faults for the early-warning systems to be effective and complex socioeconomic issues typical of developing-world communities must be addressed to reduce vulnerability.

In the United Kingdom, David King convened the Natural Hazard Working Group of scientists in 2005, and recommended the creation of an international panel to assess global natural hazards. This in turn was proposed by the United Kingdom at the G8 meeting at Gleneagles in 2005. Putting the well-received King proposals into practice has been slow. A scientific policy committee is being established within the framework of the International Strategy for Disaster Reduction (ISDR) of

resilience and saved tens of thousands of lives in Bangladesh.

A significant problem for the effective interaction of scientists and policy-makers is the rapid turnover of government officials and politicians. New governments, at least in democracies, alter policies and priorities. In many systems, civil servants and ministers frequently change posts every two or three years and the task of bringing the new people up to speed on the science must begin again.

So how might we develop policies carefully and continuously rather than in response to crises? Structures should be built that allow long-term provision of scientific advice, and analysis of hazards, risks and mitigation strategies in between disasters. The appointment of chief scientific advisers to the major government departments in the United Kingdom is an example of a step in the right direction. Governments need to ensure that the assessment activities and relevant scientific institutions are well funded, that they act on the advice with adequate funding, and that hazards remain on the agenda at all times. There is an urgent need for cooperation between social and natural scientists, as it is human activities, infrastructure and behaviour that control vulnerability and turn hazards into risks, resulting in major disasters.

Internationally, there must be more cohesive scientific input to inter-governmental policy-making. The King committee's argument for an international panel on natural hazards remains compelling. Plans for a science policy group within the ISDR are necessary, but not sufficient. There also needs to be more recognition that natural disasters are more than a development issue. Recent disasters such as this year's flooding in Britain and the wildfires in California are a reminder that natural hazards affect developed as well as developing nations.

Mitigation of natural hazards remains inadequately supported through international advisory structures. It seems that the Asian tsunami was not a strong enough wake-up call.

Steve Sparks is professor of geology at the University of Bristol, Bristol BS8 1RJ, UK.

For more essays and information see <http://nature.com/nature/focus/scipol/index.html>.



D. PARKINS



## NEWS &amp; VIEWS

## STRUCTURAL BIOLOGY

## A receptor unlocked

Stephen R. Sprang

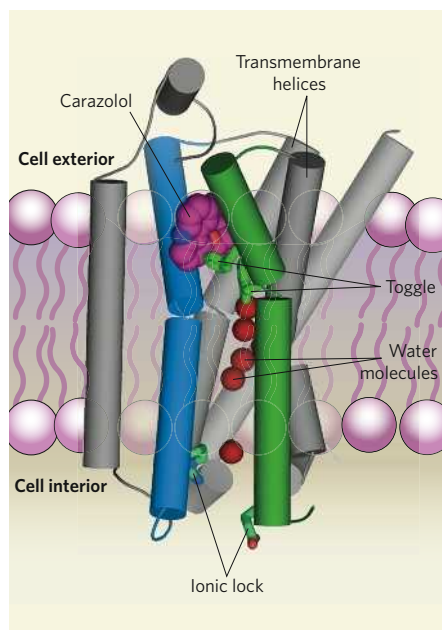
**G-protein-coupled receptors govern many biological functions, yet little is known about the molecular basis of their activity. The structure of a prominent example of these receptors is now revealed.**

The pounding heart and speeding metabolism that accompany fear and fury are the work of the adrenaline molecule, which is produced by the adrenal gland at times of stress. Adrenaline is an agonist — an activator — of  $\beta_2$ -adrenergic receptors ( $\beta_2$ ARs). Embedded in cell membranes in many tissues, these receptors control diverse effects, such as rapid pulse, pupil dilation and constriction of the blood vessels. But the  $\beta_2$ AR has even broader relevance to biologists, because it is the archetypal member of a family of more than a thousand G-protein-coupled receptors (GPCRs)<sup>1</sup>. When stimulated by agonists, GPCRs activate the eponymous G proteins that are tethered to the inner surface of the cell membrane. The G proteins in turn initiate cascades of biological signals that control such processes as gene expression, metabolic activity and cell movement.

The remarkable flexibility of GPCRs allows them to adopt several different states of activity<sup>2</sup>, but also makes it difficult to determine their three-dimensional structures. To date, only the structure of rhodopsin — a receptor involved in the perception of light, and the most stable GPCR known — has been available<sup>3</sup>. Reporting in this issue (Rasmussen *et al.*<sup>4</sup>, page 383) and in *Science* (Cherezov *et al.*<sup>5</sup> and Rosenbaum *et al.*<sup>6</sup>), several collaborating research groups now describe the structure of  $\beta_2$ AR.

Like rhodopsin,  $\beta_2$ AR consists of a protein chain that threads through the cell membrane seven times, forming a bundle of transmembrane helices (TM1–7, Fig. 1). Peptide loops connect the helices; those exposed on the outside of the cell membrane enfold the agonist binding site, whereas those on the inside of the cell form the binding site for G proteins. The G-protein binding site is formed in part from the peptide chain (known as IC3) that connects TM5 and TM6. It is thought that agonist binding induces subtle bending and twisting of the TM helices, particularly of TM3 and TM6, so exposing the G-protein binding site.

Without light to stimulate it, rhodopsin becomes completely dormant. But  $\beta_2$ ARs — in common with many other GPCRs — retain some activity in the absence of their agonists. This 'constitutive' activity can be partly



**Figure 1 | Structure of a G-protein-coupled receptor.** The cartoon shows the partial crystal structure of a modified  $\beta_2$ -adrenergic receptor ( $\beta_2$ AR), as reported by Cherezov *et al.*<sup>5</sup>. Seven helices (TM1–7) cross a cell membrane, although TM5 is not shown as it would obscure the other helices. Ligand molecules bind at a site between TM3 (blue) and TM6 (green); carazolol was used to stabilize the site in order to obtain this structure. The amino acids just below carazolol form a 'toggle' that stabilizes the inactive state of the receptor. The amino acids indicated at the bottom can form an 'ionic lock' that performs the same function, but this crystal structure and that of Rasmussen *et al.*<sup>4</sup> show that the lock is broken in  $\beta_2$ AR. Water molecules occupy a loosely packed channel; this may provide room for movement of the TM helices upon activation of the receptor, and allow binding of a G protein near the base of TM5 and TM6.

repressed by so-called inverse-agonist molecules, such as carazolol. Moreover, not all GPCR agonists are equal, as some elicit higher activity levels than others. This variable receptor output might occur because different molecules interact with the agonist binding site in different ways.

The various molecular binding modes

seem to be coupled to specific 'triggers' within GPCRs<sup>2</sup>. One such trigger is the 'ionic lock' — an electrostatic interaction between a triad of charged amino acids that tether TM3 and TM6 together. The ionic lock stabilizes the inactive state of the receptor (Fig. 1). The specific amino acids that form the lock are similar in most GPCRs; mutations that change this sequence can disrupt the lock and increase the constitutive activity of  $\beta_2$ AR. A second trigger, called the toggle, is also known, and is thought to involve a cluster of amino acids that are in contact with the agonist binding site. The toggle also seems to constrain the receptor in an inactive state; it is thought to be released when strong agonists bind to the receptor.

The molecules that bind to  $\beta_2$ AR can induce distinct conformational changes in the receptor, depending on the nature of their interactions. There is growing evidence that these different conformational states might have signalling functions in the cell other than just G-protein activation. But to understand how agonist binding couples to different conformational states, it is essential to determine the structure of the receptor at atomic resolution using X-ray crystallography.

Unfortunately, GPCRs are difficult to crystallize, because their 'greasy' hydrophobic surfaces do not readily make the regular intermolecular contacts required for crystal formation. Moreover, naturally occurring GPCRs are strongly coupled to cell membranes, which probably perturb the receptors' structures. This means that GPCRs crystallized in the absence of membranes don't necessarily have the same structure as those *in situ*. The authors of the current papers<sup>4–6</sup> all used the inverse agonist carazolol to stabilize the receptor structure, but two different methods were used to obtain a receptor complex that could be crystallized.

Rasmussen *et al.*<sup>4</sup> embedded  $\beta_2$ ARs in nanometre-sized lozenges called bicelles, which are made from bilayers of fatty molecules similar to those that constitute cell membranes. To provide a large, polar surface suitable for crystal-lattice formation, the  $\beta_2$ AR molecules were obtained as complexes with an antibody fragment that specifically binds to the IC3 loop. The antibody fragment does not interfere with

agonist binding or with the conformational changes that accompany receptor activation<sup>7</sup>. Intense, micro-focused X-ray beams produced by synchrotrons were required to tease out the diffraction data from which the crystal structures were derived.

Cherezov *et al.*<sup>5</sup> used an alternative approach. They replaced the IC3 loop of  $\beta_2$ AR with a small, stable protein known as T4 phage lysozyme (T4L). The T4L protein promotes crystal-lattice formation in the same way as the antibody fragment used by Rasmussen *et al.*<sup>4</sup>. The modified receptor, bound to carazolol, was crystallized in a semi-solid lipid medium that provided an artificial, membrane-like environment for the receptor. Synchrotron X-rays were also required in this case to obtain good-quality diffraction data. The results of these investigations are described by Rosenbaum *et al.*<sup>6</sup>.

A surprise finding from the crystal structures<sup>4,5</sup> is that the ionic lock is broken (as expected in an active state) despite the presence of an inverse agonist (which promotes the inactive state). Had only one crystal form of  $\beta_2$ AR been available, it could have been argued that this was an artefact of the experimental conditions — either a consequence of the artificial lipid environments, or because of perturbation of IC3. But because the lipid supports and IC3 region are different in each structure, it is unlikely that the same misleading result could have occurred in both cases. Moreover, the TM regions in both structures align well with each other, further corroborating the results.

Thus, the weak constitutive activity of  $\beta_2$ AR might be attributed to the breaking of the ionic lock; this can be compared with rhodopsin, which retains the lock and shows no such activity. In contrast, the toggle is intact in the two  $\beta_2$ AR structures, just as it is in inactivated rhodopsin, despite considerable differences in the binding modes adopted by the ligand molecules for these two GPCRs. Perhaps stronger inverse agonists than carazolol would stabilize the ionic lock and fully inactivate  $\beta_2$ AR, whereas agonists would release both lock and toggle. But these speculations have yet to be confirmed.

The crystal structures<sup>4,5</sup> reveal several other interesting clues to  $\beta_2$ AR behaviour. For example, many of the amino acids that produce constitutive activity are known to respond to agonist binding, but are not directly connected to the agonist binding site. It can now be seen that these residues are linked to the binding site via packing interactions<sup>6</sup>, so that movement of amino acids in the binding site could affect the packing of others throughout the structure (Fig. 1). Furthermore, a water-filled channel in the core of the receptor offers space for structural rearrangement, which could enable the receptor to adopt several different active states. Finally, the capacious agonist binding site is compatible with the existence of several molecular binding modes, each with the potential to trigger a different state in the receptor.

To learn how structural changes are conveyed

from the agonist pocket to the G-protein binding site, the structure of the fully active, agonist-bound  $\beta_2$ AR must be determined. The instability of the activated receptor will make this difficult — in fact, it may be possible to resolve the structure of a receptor only if it is bound to a G protein. Nevertheless, with the structures<sup>4–6</sup> of  $\beta_2$ AR in hand, we can expect those of other GPCRs to follow, so that the conformational complexities common to the members of this family of receptors might finally be revealed. ■

Stephen R. Sprang is at the Center for Biomolecular

Structure and Dynamics, University of Montana, Missoula, Montana 59812, USA.

e-mail: stephen.sprang@umontana.edu

1. Pierce, K. L., Premont, R. T. & Lefkowitz, R. J. *Nature Rev. Mol. Cell Biol.* **3**, 639–650 (2002).
2. Kobilka, B. K. & Deupi, X. *Trends Pharmacol. Sci.* **28**, 397–406 (2007).
3. Palczewski, K. *et al. Science* **289**, 739–745 (2000).
4. Rasmussen, S. G. F. *et al. Nature* **450**, 383–387 (2007).
5. Cherezov, V. *et al. Science* doi:10.1126/science.1150577 (2007).
6. Rosenbaum, D. M. *et al. Science* doi:10.1126/science.1150609 (2007).
7. Day, P. W. *et al. Nature Meth.* doi:10.1038/nmeth1112 (2007).

## PLANETARY SCIENCE

# Isotopic lunacy

Alex N. Halliday

**The Moon could have been derived from a well-mixed disk of rock vapour that was produced after the early Earth collided with another planet. This persuasive idea offers a fresh perspective on the history of both bodies.**

“It is the very error of the moon; She comes more nearer earth than she was wont; And makes men mad.” Thus spoke Othello, who was talking of a murderous madness. For 30 years scientists have been gripped, albeit with less deadly consequences, by a maddening paradox to do with the Moon: how near its oxygen-isotope composition is to that of Earth.

It is generally agreed that the Moon formed in a giant impact between Earth and another, smaller body. But previous simulations of that event show that, dynamically, the Moon should be largely derived from the smaller body, in which case its oxygen isotopic composition should be different from that of Earth. Writing in *Earth and Planetary Science Letters*, Pahlevan and Stevenson<sup>1</sup> put forward an explanation for the oxygen-isotope paradox: they propose that the giant impact produced a disk of rock vapour within which the atoms were able to mix before the lunar component condensed out.

The scientific context of this story goes back to the discovery of large variations in the isotopic composition of oxygen in meteorites<sup>2</sup>, which occurred at about the same time as lunar samples were being returned by Apollo astronauts. These variations are so large that they can readily be used to identify which meteorites come from the same planet or asteroid. Lunar samples were found to have a terrestrial isotopic composition<sup>2</sup>, which was taken as evidence that the Moon was derived from Earth<sup>3</sup>. But the evidence from dynamic simulations is that most of the mass of the Moon came from another planet, Theia, that hit the early Earth with a glancing blow<sup>4</sup>.

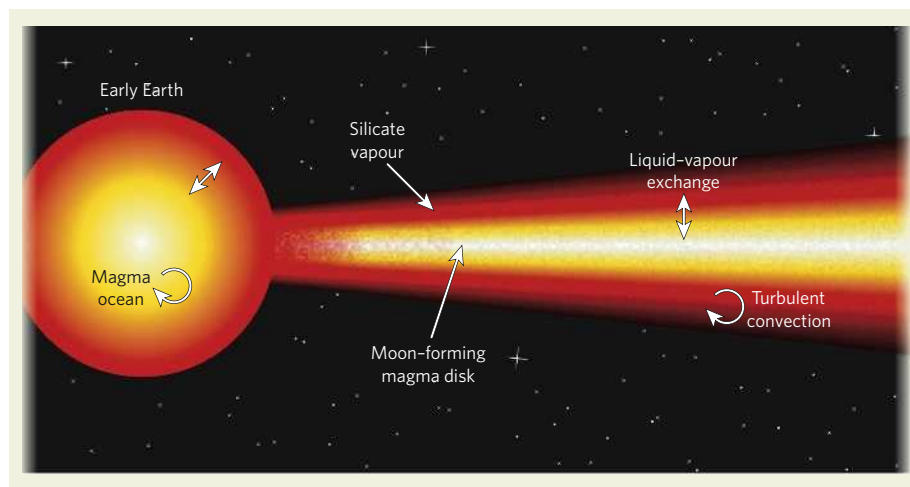
The isotopic composition of every element studied so far is the same in lunar samples as it is in samples from Earth, apart from elements that would change because of radioactive

decay, or through the effects of cosmic rays or the charged particles of the solar wind. But the background Solar System heterogeneity is far bigger for oxygen than for most other elements<sup>2</sup>. So, with the development of more precise techniques, a further analysis of lunar samples aimed to resolve an Earth–Moon oxygen-isotope difference that it was thought must surely exist at some finite level<sup>5</sup>. Far from doing that, however, samples from the two bodies were found to be identical to within better than five parts per million. One explanation was that Theia must have formed from the same mix of material as Earth.

Pahlevan and Stevenson’s dynamic simulations<sup>1</sup> of Solar System formation build on previous work in assuming that, based on the measured difference between Earth and Mars, an oxygen-isotopic gradient existed within the early planetesimals of the circumstellar disk. This is a reasonable assumption, but might not be easy to test at the required precision until samples are returned from sizeable bodies such as Mercury and Venus. The obvious test at present involves a class of meteorites that were probably derived from the asteroid 4 Vesta. The oxygen-isotope data for these meteorites do not fit with the proposed radial-gradient model. But this can be explained if, as seems likely, Vesta has migrated outwards from the innermost Solar System.

Based on this putative gradient, the new calculations<sup>1</sup> show that it is unlikely for two bodies such as Theia and Earth to have had the same bulk isotopic composition. Indeed, it is hard to imagine how they could have formed from the same mix of primordial matter unless they grew at exactly the same distance from the Sun, which in turn raises the question of why it took them so long to collide. Tungsten-





**Figure 1 | The big mix-up.** According to Pahlevan and Stevenson's model<sup>1</sup>, the immense energy released in the collision between the early Earth and another body vaporized Earth's outer, silicate portion and created an internal ocean of liquid magma. The material from which the Moon eventually formed was a disk of magma created in the impact, and derived from the impacting body, which was connected to Earth by a shared atmosphere of silicate vapour. Through the processes of turbulent convection and exchanges between the different phases driven by heat loss, matter from the two bodies became well mixed on a timescale of  $10^2$ – $10^3$  years. This scheme can account for the fact that although the oxygen-isotope content of both the proto-Earth and the impactor were probably dissimilar, those of today's Earth and Moon are virtually identical. (Derived from Fig. 3 of ref. 1.)

isotope measurements indicate that the Moon did not form until about 4.5 billion years ago, more than 40 million years after the birth of the Solar System<sup>6</sup>.

The energy released when sizeable worlds collide is astonishing by geological standards. The giant impact would almost certainly have melted most of the Earth and produced a magma ocean<sup>7</sup>. Much of the Earth and Theia may have been vaporized. Pahlevan and Stevenson<sup>1</sup> calculate that over the time required for the disk of vapour to cool and condense, the atoms of the silicate portion of Earth — that is, everything but the metal core — would have been vigorously and repeatedly stirred, mixing with the atoms of the disk and eliminating isotopic variations (Fig. 1).

Support for Pahlevan and Stevenson's theory comes from a study<sup>8</sup> of silicon isotopic compositions, which differ in meteorites and samples from the silicate Earth. The latter are fractionated to heavier isotopic compositions, and this is thought to have been caused by equilibration with silicon that was incorporated into Earth's core under very high pressures and temperatures. Theia was a small (Mars-sized) planet that should have had a silicon isotopic composition like that of meteorites and Mars. However, the Moon's composition is heavy, like that of the silicate Earth<sup>8</sup> — which is hard to explain unless Pahlevan and Stevenson are correct and there was large-scale equilibration.

If they are indeed right, it means that certain features of the Moon's composition reflect those of the early Earth. So, since the first lunar samples were returned, we have, without realizing it, been analysing a unique archive about our own planet — a body that has lost almost all traces of its early development with 4.5 billion years of subsequent bombardment, mantle

convection and geological reprocessing.

The new theory raises lots of questions. For example, it is unclear whether all isotopic systems would have been efficiently mixed. Mixing is related to volatilization, so highly refractory elements might still show differences. Exploring this possibility will probably require improvements in mass spectrometry to resolve tiny isotopic differences — unlike oxygen, most elements in the periodic table have an almost identical proportion of isotopes, whether measured on a lump of the Empire State Building or a meteorite from the asteroid belt.

Another issue is that, although the Moon is isotopically identical to Earth, it is chemically

different. In particular, it is far more depleted in volatile elements such as alkalis, even though these elements should have mixed especially efficiently. This can be explained if volatiles were somehow lost from the hot disk from which the Moon condensed. But other differences are harder to account for. The iron content of lunar basalt rocks is higher than that of Earth's and more like those of Mars' and Vesta's. This is thought to reflect a higher iron content in the lunar mantle. The lower iron content of Earth's mantle was thought to reflect an early, less-oxidized stage in the growth of the Earth<sup>9</sup>. However, if the Moon is essentially a sample of Earth at the time of the giant impact, it may indicate that Earth's mantle originally had a higher iron content that was subsequently depleted by formation of the core.

The ramifications of Pahlevan and Stevenson's model<sup>1</sup> extend well beyond these issues. Testing and developing this exciting new theory will be priorities over the coming years — the maddening paradox may well be about to generate a new wave of highly creative science.

Alex N. Halliday is in the Mathematical, Physical and Life Sciences Division, University of Oxford, 9 Parks Road, Oxford OX1 3PD, UK.  
e-mail: alex.halliday@mpls.ox.ac.uk

1. Pahlevan, K. & Stevenson, D. J. *Earth Planet. Sci. Lett.* **262**, 438–449 (2007).
2. Clayton, R. N. in *Stable Isotopes in High Temperature Geological Processes* (eds Valley, J. W., Taylor, H. P. & O'Neill, J. R.) 129–140 (Mineral. Soc. Am., Washington DC, 1986).
3. Ringwood, A. E. *Earth Planet. Sci. Lett.* **95**, 208–214 (1989).
4. Canup, R. M. & Asphaug, E. *Nature* **412**, 708–712 (2001).
5. Wiechert, U. *et al. Science* **294**, 345–348 (2001).
6. Halliday, A. N. *Nature* **427**, 505–509 (2004).
7. Solomatov, V. S. in *Origin of the Earth and Moon* (eds Canup, R. M. & Righter, K.) 323–360 (Univ. Arizona Press, Tucson, 2000).
8. Georg, R. B., Halliday, A. N., Schauble, E. A. & Reynolds, B. C. *Nature* **447**, 1102–1106 (2007).
9. Wade, J. & Wood, B. J. *Earth Planet. Sci. Lett.* **236**, 78–95 (2005).

## EPIGENETICS

# Reversing the 'irreversible'

Richard S. Jones

**"Do not speak — unless it improves on silence" is generally wise advice, and is even vital for a subset of essential genes. New studies describe how, when appropriate, the silence of these genes is broken.**

In cellular tissues, it is essential that certain genes are turned on in appropriate cells but remain silent in others. For years, the dogma has been that some forms of gene silencing are irreversible, or at least extremely stable. One such form of silencing is mediated by the Polycomb group of proteins, which repress gene expression partly by altering the structure of chromatin — complexes of DNA and histone proteins. Such repression can be achieved by,

for example, adding a specific molecular tag to a histone. Five papers<sup>1–5</sup>, including one by Jepsen *et al.*<sup>1</sup> on page 415 of this issue, now describe how enzymatic removal of one molecular tag reverses the silencing of gene expression that is mediated by the Polycomb group.

It has long been known that the way genes are packaged inside the nucleus of a cell can determine their expression (transcription). Molecular tags on histones have crucial roles

in controlling the activity of associated genes; such tags include acetyl, methyl and phosphate groups. Each of these tags may be added to or removed from specific amino acids in histones. The particular array of tags on histones at a specific genomic location may either facilitate or hinder the binding of regulatory proteins to DNA, resulting in the activation or inhibition of transcription<sup>6</sup>.

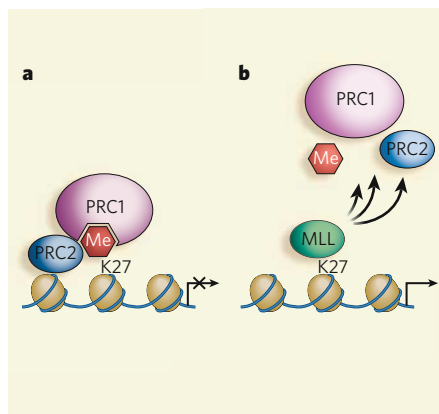
Of the various types of histone modification, methylation is the most stable. For example, precise patterns of trimethylated lysine residues at positions 9 or 27 of the amino-acid sequence of histone H3 can be maintained through many cell cycles in specific cell lineages, preserving the silenced state of the associated genes. However, in recent years, several enzyme groups that remove methyl tags from histones have been identified. The largest group of these histone demethylase enzymes is characterized by a catalytic structural domain called JmjC (refs 7, 8). Each of these JmjC-containing enzymes removes a methyl group from a specific lysine residue of histone H3.

The only enzymes that methylate lysine 27 of histone H3 are members of the Polycomb-group of transcriptional repressors<sup>9</sup>. The mammalian methylase enzyme for this reaction is EZH2, which functions as part of a protein complex called PRC2, adding a trimethyl group to lysine 27. This leads to the recruitment of other Polycomb-group proteins and silencing of target genes (Fig. 1a). These include genes that are essential for regulating developmental pathways, such as *HOX* genes.

The reversible nature of Polycomb-group-mediated gene silencing and the significance of this regulatory plasticity for development and other physiological processes have become apparent only recently<sup>9</sup>. But because no demethylase enzyme had been identified for trimethylated lysine 27 of histone H3, there was a growing sense that Polycomb-group-mediated gene silencing might be reversed only by the slow loss of the trimethyl group from this lysine residue during cell proliferation, following removal of the PRC2 protein complex from target genes.

The new studies<sup>1–5</sup> change this view, demonstrating that the trimethylated lysine is specifically demethylated by two similar JmjC-containing enzymes — JMJD3 and UTX. Both of these demethylases are components of MLL protein complexes, which antagonize Polycomb-group-mediated gene silencing (Fig. 1b).

The *SMRT* gene is known to be associated with neuronal development, and Jepsen *et al.*<sup>1</sup> studied the brains of mouse embryos in which *SMRT* was deleted and also cultured cells derived from them. They found that the product of this gene is needed to maintain neural stem cells in their undifferentiated state. They also found that *SMRT*-deficient progenitor cells have a strong tendency to differentiate either into neurons or into non-neuronal



**Figure 1 | Reversal of Polycomb-group silencing by demethylases.** DNA (blue) wraps around histone proteins (brown spheres) to form nucleosomes, the building-blocks of chromatin. **a**, In the silenced state, the PRC2 complex, which contains Polycomb-group proteins, binds to target genes and adds three methyl groups (Me) to histone H3 at lysine 27 (K27). A second Polycomb-group-containing complex, PRC1, is recruited, partly because a component of this complex has a specific affinity for histone H3 trimethylated at K27. Consequently, PRC1, and possibly other Polycomb-group proteins, prevent transcription of the genes associated with this lysine. **b**, Five studies<sup>1–5</sup> suggest that such gene silencing might be rapidly reversed by the MLL class of protein complexes, which contain either of two histone demethylase enzymes, JMJD3 and UTX. The binding of MLL complexes to regions containing histone H3 trimethylated at K27 is associated with displacement of PRC2, allowing JMJD3 or UTX to actively demethylate this particular lysine. Consequently, the association between PRC1 and the target gene is destabilized, permitting gene transcription and additional modifications that also promote transcription.

brain support cells called glia. The authors explain these observations in terms of a direct, inhibitory effect of the SMRT protein on the expression of JMJD3.

The SMRT protein is a co-repressor that works with a nuclear receptor known as the retinoic-acid receptor to prevent expression of retinoic-acid-responsive genes. Treatment of normal cells with retinoic acid results in the dissociation of SMRT from the retinoic-acid receptor, triggering JMJD3 expression. Jepsen *et al.* suggest that, in turn, JMJD3 removes the trimethyl group from lysine 27 of histone H3, thereby contributing to the de-repression of several genes that trigger neurogenesis. This seems to happen while the levels of the EZH2 methylase enzyme remain constant.

Through this model, Jepsen and colleagues propose a mechanism by which histone tags added by Polycomb-group proteins are rapidly removed in response to retinoic-acid signalling. Consequently, neural stem cells are released to take on more specialized roles.

Two other studies<sup>2,3</sup> report that retinoic-acid signalling leads to the recruitment of another

enzyme that demethylates lysine 27 — UTX. This enzyme acts on multiple *HOX* genes in cultured human embryonic teratocarcinoma cells, causing displacement of EZH2, demethylation of lysine 27 and de-repression of the *HOX* genes. When mouse embryonic stem cells are treated with retinoic acid, EZH2 is rapidly displaced from retinoic-acid-responsive genes and the histone H3 lysine 27 residues associated with these genes are swiftly demethylated. Intriguingly, these changes are much slower in genes that are not retinoic-acid responsive<sup>10</sup>. This suggests a dual mode of ‘reawakening’ Polycomb-group-silenced genes in stem cells: a fast mode of de-repression, possibly involving demethylation of lysine 27 by JMJD3 or UTX; and a slow mode resulting from passive diminution of trimethylated lysine 27 in the absence of PRC2.

A further study<sup>4</sup> finds that the JMJD3 demethylase also participates in regulating the differentiation of bone-marrow cells and the activation of immune-system cells called macrophages. When macrophages are exposed to bacterial products or other activators, JMJD3 expression is induced by a transcription factor called NF- $\kappa$ B, facilitating the activation of a Polycomb-group-silenced gene, *BMP-2*. So both the retinoic-acid receptor and NF- $\kappa$ B can trigger JMJD3 expression, suggesting that JMJD3 is regulated by different external stimuli in different cell types. In both cases, however, the targets of JMJD3 are discrete sets of genes that were initially silenced by Polycomb-group proteins.

Clearly, the identification of two demethylases for lysine 27 of histone H3 is only the first step in understanding their antagonistic relationship with Polycomb-group silencing proteins. Many questions remain concerning targeting of these enzymes to specific genes, their roles in the initial establishment of patterns of trimethylated lysine 27, and other factors with which they might interact to initiate transcription. Moreover, genetic analysis of JMJD3, UTX and their homologues in other species will probably reveal other developmental and physiological processes that depend on the reversal of Polycomb-group-mediated gene silencing. ■

Richard S. Jones is in the Department of Biological Sciences, Southern Methodist University, Dallas, Texas 75275-0376, USA. e-mail: rjones@smu.edu

1. Jepsen, K. *et al.* *Nature* **450**, 415–419 (2007).
2. Agger, K. *et al.* *Nature* **449**, 731–734 (2007).
3. Lee, M. G. *et al.* *Science* **318**, 447–450 (2007).
4. De Santa, F. *et al.* *Cell* **130**, 1083–1094 (2007).
5. Lan, F. *et al.* *Nature* **449**, 689–694 (2007).
6. Margueron, R., Trojer, P. & Reinberg, D. *Curr. Opin. Genet. Dev.* **15**, 163–176 (2005).
7. Klose, R. J. & Zhang, Y. *Nature Rev. Mol. Cell Biol.* **8**, 307–318 (2007).
8. Shi, Y. & Whetstone, J. R. *Mol. Cell* **25**, 1–14 (2007).
9. Sparmann, A. & van Lohuizen, M. *Nature Rev. Cancer* **6**, 846–856 (2006).
10. Lee, E. R., Murdoch, F. E. & Fritsch, M. K. *Stem Cells* **25**, 2191–2199 (2007).



## CARBON CYCLE

# Quick burial at sea

Caroline A. Masiello

**The amount of river-borne carbon that is buried upon reaching the sea affects Earth's atmospheric composition. A study of rivers draining the Himalaya shows that carbon burial may occur more efficiently than was thought.**

It's a popular misconception that the concentration of oxygen in Earth's atmosphere is controlled by photosynthesis. Photosynthesis is certainly the source of atmospheric  $O_2$ , but the amount it produces is in almost perfect balance with the amount consumed through the respiration of living organisms. It is only when organic matter is buried in ocean sediments, and so ceases to be decomposed, that atmospheric  $O_2$  can accumulate. This burial process also reduces the levels of the greenhouse gas carbon dioxide released into the atmosphere. The exact rate of organic-matter burial is therefore a significant determinant of atmospheric composition, and thus global climate, over geological timescales.

On page 407 of this issue, Galy *et al.*<sup>1</sup> bring a new perspective on how organic carbon is stored in sediments. They show that the Ganges–Brahmaputra river system, which drains the Himalaya, interacts with the sediments deposited by the rivers' waters in the Bay of Bengal — the 'Bengal fan' — so as to store organic carbon more efficiently than studies on other river systems had suggested would be the case.

The fact that organic matter can be stored at all in ocean sediments is something of a surprise. Heterotrophic organisms (those that acquire their carbon from other organic matter, rather than synthesizing it themselves from inorganic sources) operate highly efficiently in the oceans, tightly recycling carbon and nutrients in the ocean's surface layers. Less than 1% of photosynthesized organic matter makes it down to the surface of ocean sediments<sup>2</sup>, and what carbon does hit the seafloor enters a voracious ecosystem that makes short work of almost all of it<sup>3</sup>. The little burial of organic matter that does take place occurs in river deltas and at continental margins. In addition, rivers deliver more than enough carbon to the oceans to account for the total amount of carbon stored in sediments<sup>4</sup>. These two facts have long raised suspicions that terrestrially derived organic matter is a significant source of the carbon preserved in marine sediments.

But data from rivers and coastal sediments have created a more complex picture. In particular, measurements made in the world's largest river system, the Amazon, have shown that even the strongest bonds between minerals and organic matter formed in soils can be broken during transport and deposition at the river–ocean margin. This causes terrestrial carbon in mineral surfaces to be replaced, at least

in part, by marine carbon<sup>5</sup>. Indeed, the fan of sediment spreading out from the mouths of the Amazon has been called a "gigantic sedimentary incinerator"<sup>6</sup> for its efficiency at destroying both marine and terrestrial organic matter, preventing their burial at sea. There had been hints that other types of river — in particular, mountainous rivers on active tectonic margins such as those draining the Himalaya — export and store carbon differently from the Amazon<sup>7,8</sup>. But these rivers are individually small, making it difficult to place their stories in a globally significant context.

In their studies of the Ganges–Brahmaputra system, Galy *et al.*<sup>1</sup> use the relationship between organic-carbon content and the ratio of aluminium to silicon in river and Bengal Fan sediments as a marker for the amount of organic matter these particles contain. They find that this relationship remains the same from the rivers' youthful stages right through to sediment deposition in the ocean (Fig. 1). Coupling this information with earlier studies using sedimentary biomarkers that showed no evidence of the storage of marine organic carbon in Bengal-fan sediments<sup>9</sup>, Galy *et al.*<sup>1</sup> conclude, reasonably, that the terrestrial

organic matter is stored in the Bengal fan very efficiently — in contrast to the picture of destruction at the mouths of the Amazon.

The authors attribute this efficiency to the Bengal fan's rapid sedimentation rate, which restricts the sediments' exposure to oxygen, effectively cutting off microbial decomposition of organic matter. They further argue that the driver for the rapid sedimentation is the fast erosion resulting from tectonic uplift in the Himalaya, an argument that meshes with hypotheses previously published for much smaller rivers draining tectonically active margins<sup>7</sup>.

But is this all there is to it? Those of us who study the river transport of organic matter to the ocean make the unstated assumption that carbon loading does not vary with the soil properties of the particular river basin. We assume that rocks break apart into soil minerals, soil minerals load with organic matter, and organically coated minerals are exported to rivers by essentially the same mechanisms the world over, with soil type and degree of development influencing the final storage outcome only insignificantly. Yet soil scientists know that soil mineralogy has an important part in determining the mechanisms and durability of organo-mineral bonds: for example, the effectiveness of carbon storage in soil varies greatly with the degree of soil weathering<sup>10</sup>.

The Amazon and Ganges–Brahmaputra river systems are ideal for testing the effect of river-basin soils on the storage of organic carbon in delta sediments, because they drain environments that represent two extremes of soil development. The Amazon basin is dominated by Oxisols, the most highly weathered



**Figure 1 | From mountains to the sea.** In the Ganges–Brahmaputra river system, particles eroded from the Himalaya weather and become loaded with organic carbon in the ecosystems of the drainage basin. With their coatings of terrestrial carbon intact, these particles are delivered to the ocean floor off the coast of Bangladesh. Galy and colleagues' study<sup>1</sup> indicates that the carbon content of these sediments is efficiently preserved in the Bengal sedimentary fan, preventing this carbon from being released to the atmosphere as  $CO_2$ .

J. DESCLOITRES, MODIS LAND/RAPID RESPONSE TEAM AT NASA GSFC

type of soil, whereas the Ganges–Brahmaputra basin is dominated by Inceptisols, one of the least weathered soil orders. Understanding the influence of river-basin soils on the storage of carbon in ocean sediments will require the effective integration of soil data into the field of oceanography. Carbon-coated minerals travel from mountains through ecosystems to rivers and ocean sediments: perhaps understanding these particles' history will help us predict their sedimentary fate. ■

Caroline A. Masiello is in the Department of Earth Science, Rice University, Mail Stop 126, 6100 Main Street, Houston, Texas 77005, USA. e-mail: masiello@rice.edu

1. Galy, V. *et al.* *Nature* **450**, 407–410 (2007).
2. Hedges, J. I. & Keil, R. G. *Mar. Chem.* **49**, 81–115 (1995).
3. Witte, U., Aberle, N., Sand, M. & Wenzhofer, F. *Mar. Ecol. Prog. Ser.* **251**, 27–36 (2003).
4. Hedges, J. I., Keil, R. G. & Benner, R. *Org. Geochem.* **27**, 195–212 (1997).
5. Keil, R. G., Mayer, L. M., Quay, P. D., Richey, J. E. & Hedges, J. I. *Geochim. Cosmochim. Acta* **61**, 1507–1511 (1997).
6. Aller, R. C. & Blair, N. E. *Cont. Shelf Res.* **26**, 2241–2259 (2006).
7. Milliman, J. D. & Syvitski, J. P. M. *J. Geol.* **100**, 525–544 (1992).
8. Blair, N. E., Leithold, E. L. & Aller, R. C. *Mar. Chem.* **92**, 141–156 (2004).
9. Freeman, K. H. & Colarusso, L. A. *Geochim. Cosmochim. Acta* **65**, 1439–1454 (2001).
10. Torn, M. S., Trumbore, S. E., Chadwick, O. A., Vitousek, P. M. & Hendricks, D. M. *Nature* **389**, 170–173 (1997).

## MICROBIOLOGY

# Pathogen drop-kick

Suzanne Pfeffer

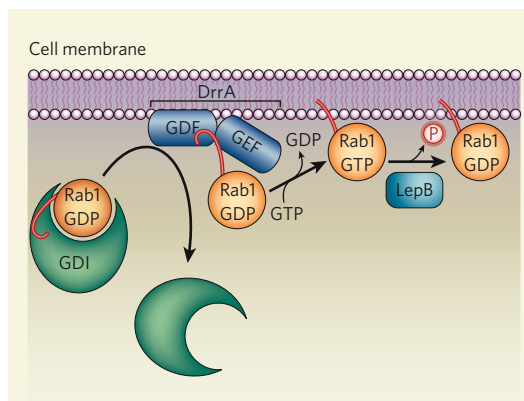
**To set the scene for its replication, the bacterium *Legionella pneumophila* exploits its host cells' Rab1 protein. This pathogen seems to use minimal resources to mimic the normal cycle of Rab1 activity.**

The bacterium *Legionella pneumophila* is the main agent responsible for both Legionnaires' disease, which leads to pneumonia, and Pontiac fever, an illness that resembles acute influenza. This pathogen is engulfed by immune cells known as macrophages and, once inside, exploits proteins mediating intracellular trafficking to form an organelle called an *L. pneumophila*-containing vacuole, or LCV. The bacterium then replicates inside the LCV. Two papers<sup>1,2</sup>, including one by Ingmundson and colleagues that appears on page 365 of this issue, provide insight into how *L. pneumophila* uses one host-cell protein to form LCVs.

Among the proteins that mediate intracellular trafficking are Rab proteins; the human genome encodes about 70 of these proteins<sup>3</sup>. Rab proteins are involved in organizing membranes for the formation of vesicular carriers, the transport of these vesicles between cellular compartments and vesicle fusion with the membrane of target cellular compartments. Rab proteins interconvert between an active, GTP-bound state and an inactive, GDP-bound form. To activate a Rab protein, enzymes known as GEFs exchange its bound GDP nucleotide for a GTP, whereas for Rab inactivation, enzymes known as GAPs help Rabs hydrolyse Rab-bound GTP to produce GDP.

So the Rab cycle of activity goes as follows. An active, GTP-bound Rab uses two 20-carbon-long 'anchors' to bind to the membrane surface of distinct organelles. There, it organizes the formation of functional membrane

microdomains. GAP-stimulated hydrolysis of Rab's GTP to GDP follows, allowing a cytoplasmic protein called GDI to bind to the inactive Rab and extract it from membranes<sup>4</sup>. Afterwards, another protein, GDI displacement factor, or GDF, releases Rab from GDI, and GEF enzymes then mediate the exchange of Rab's GDP molecule for a GTP, preparing Rab for its next cycle of activity<sup>5,6</sup>.



**Figure 1 | How *Legionella pneumophila* exploits host-cell Rab1 protein.** In the host-cell cytoplasm, the inactive, GDP-bearing Rab1 associates with GDI. This complex is recognized by the DrrA protein of *L. pneumophila*, which localizes to the inner leaflet of the cell membrane. Two studies<sup>1,2</sup> find that one region of DrrA acts as a GDF, releasing Rab1 from GDI, whereas another domain of this protein functions as a Rab1-specific GEF to activate Rab1 by catalysing the exchange of its associated GDP for a GTP. The active Rab1 can then anchor into the adjacent membrane to organize the formation of a functional membrane microdomain. Later, another *L. pneumophila* protein, LepB, stimulates the hydrolysis of Rab1-bound GTP to GDP and inorganic phosphate (P), thereby inactivating Rab1.

Ingmundson *et al.*<sup>1</sup> and the authors of a recent *Science* paper<sup>2</sup> demonstrate that, to harness the host-cell cycle of Rab activity, an *L. pneumophila* protein known as DrrA functions as both a GDF and a GEF. This pathogen specifically targets Rab1, which normally mediates vesicle transport from the endoplasmic reticulum to the Golgi complex.

Previous work<sup>7,8</sup> had shown that DrrA can both recruit Rab1 to LCVs and activate it. Furthermore, the carboxyl terminus of DrrA (that is, amino-acid residues 451–647) has been shown to function as a Rab1-specific GEF *in vitro*, explaining the ability of this protein to activate Rab1 (ref. 7). In living cells, however, this region of DrrA was insufficient to recruit Rab1 to LCVs.

These observations led the authors<sup>1,2</sup> to test whether DrrA also functions as a GDF to release Rab1 from GDI and recruit it to LCVs. Ingmundson *et al.* find<sup>1</sup> that amino-acid residues 1–500 of DrrA — which lack GEF activity — can efficiently release Rab1 from GDI. This suggests that two separate regions within a single *L. pneumophila* protein contain GDF and GEF activities, allowing it to drop-kick Rab1 off GDI and onto adjacent membranes in an active form (Fig. 1).

Structurally, such an arrangement makes sense. The so-called 'switch' domains that participate in GTP/GDP nucleotide binding are the only parts of Rab proteins to change conformation between the GDP- and GTP-bound states. So proteins such as GDI that specifically interact with the GDP-bearing form of the 'switchable' proteins must recognize the status of the switch domains; the crystal structure<sup>9</sup> of a Rab bound to GDI shows that this is the case. Similarly, to catalyse nucleotide exchange, a GEF must also access the switch domains. Thus, with GDI tightly bound to Rab proteins, GDF must move the GDI off the switch domain to permit a GEF to do its job.

A GDF can localize a Rab protein to a specific membrane after dissociating it from a GDI (ref. 10). DrrA is found on the inner leaflet of the infected host-cell membrane and, in cells that express DrrA exogenously, this protein can mis-target Rab1 so that it also localizes to the cell membrane<sup>7,8</sup>. It will be interesting to determine how DrrA localizes to the inner cell membrane in the first place.

Ingmundson *et al.*<sup>1</sup> also show that another *L. pneumophila* protein, LepB, is a Rab1-specific GAP enzyme. Thus, these bacteria seem to be self-sufficient in conducting the Rab1 cycle: they build a functional membrane microdomain by recruiting and locally activating Rab1, and, when appropriate, disassemble this microdomain using a bacterially encoded GAP (Fig. 1). The bacteria eventually return Rab1 to host-cell service, which might in itself be beneficial to a productive infection process.

What are the main functions of Rab1





## 50 YEARS AGO

There can be few people now who are not convinced that the British public schools are aware of the country's need of scientific man-power. Not only is the teaching of science in these schools thoroughly 'respectable' now, but also the numbers specializing in science in the sixth form are often more than 50 per cent of the total numbers ... The task now is to give every boy his chance to train as a scientist if he wishes ... It must be realized, too, that for every pure scientist we need many technologists, and for every technologist we need an army of technicians.

From *Nature* 16 November 1957.

## 100 YEARS AGO

I see in *NATURE* of November 7 advertisements for a principal and professor of chemistry for the Victoria Technical College in Bombay. Though I have no connection with the institute, and may be charged with unwarrantable interference, I think that it is only fair to the intending competitors that certain facts should be made known ... The title of principal does not even secure to the holder of it the exclusive right of calling meetings of the staff, and, in fact, confers nothing more than the power to enforce discipline ... Such were the conditions under which my friend Dr. Mackenzie held the appointment, and it was with no astonishment that I heard on my return from India that he had sent in his resignation. I may add that though Rs. 1000 per mensem with a residence appears to be a good salary, it must be remembered that there is no security of tenure of the appointment, and that the residence offered to Dr. Mackenzie lay between the dustiest road and the busiest railway in the heart of Bombay. Should any chemist contemplate applying for the "chair" of chemistry with the view of carrying out research ... I should like to remind him that he will do well to take his library with him. There are no scientific books in Bombay...

Morris W. Travers

From *Nature* 14 November 1907.

on the LCV? What specific proteins does it interact with there, and how do they benefit bacterial infection? Are there any proteins encoded by the host cell that also have dual GDF/GEF activity? Answers to these questions may lead to selective approaches for treating *L. pneumophila* infections, and a better understanding of the functioning of Rab proteins. ■ Suzanne Pfeffer is in the Department of Biochemistry, 279 Campus Drive, Room B400, Stanford University School of Medicine, Stanford, California 94305-5307, USA. e-mail: pfeffer@stanford.edu

1. Ingmundson, A., Delprato, A., Lambright, D. G. & Roy, C. R. *Nature* **450**, 365–369 (2007).
2. Machner, M. P. & Isberg, R. R. *Science* doi:10.1126/science.1149121 (2007).
3. Grosshans, B. L., Ortiz, D. & Novick, P. *Proc. Natl Acad. Sci. USA* **103**, 11821–11827 (2006).
4. Sasaki, T. et al. *J. Biol. Chem.* **265**, 2333–2337 (1990).
5. Soldati, T., Shapiro, A. D., Dirac Svejstrup, A. B. & Pfeffer, S. R. *Nature* **369**, 76–78 (1994).
6. Sivars, U., Aivazian, D. & Pfeffer, S. R. *Nature* **425**, 856–859 (2003).
7. Murata, T. et al. *Nature Cell Biol.* **8**, 971–977 (2006).
8. Machner, M. P. & Isberg, R. R. *Dev. Cell* **11**, 47–56 (2006).
9. Pylypenko, O. et al. *EMBO J.* **25**, 13–23 (2006).
10. Pfeffer, S. & Aivazian, D. *Nature Rev. Mol. Cell Biol.* **5**, 886–896 (2004).

## QUANTUM OPTICS

# Kittens catch phase

Jonathan P. Dowling

**The most sensitive phase-measuring instrument yet uses quantum trickery and a scaled-down version of the notorious Schrödinger's cat. It means that more sensitive devices for metrology and imaging could be on the way.**

Elsewhere in this issue, a group of Australian researchers (Higgins *et al.*, page 393)<sup>1</sup> reports the construction of the most sensitive optical interferometer yet. At the heart of this success lies a cunning use of quantum feedback to minimize photon phase noise in the device — a technique that could have applications in imaging, remote sensing, gravity-wave detection and spectroscopy.

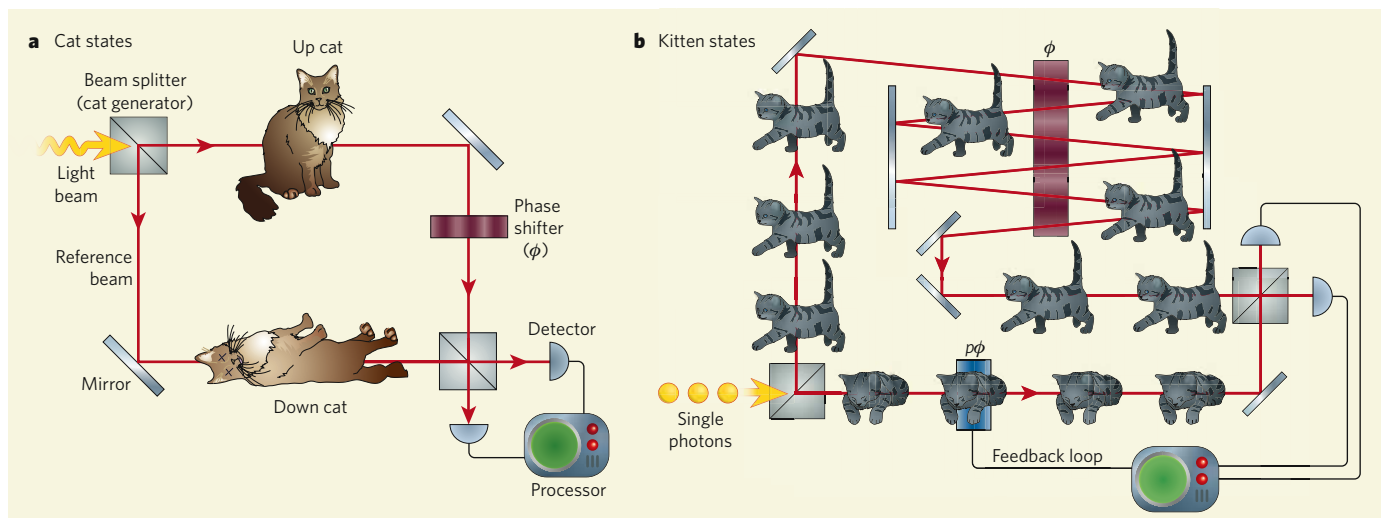
An interferometer detects differences in the phase of two waves by measuring the extent to which their amplitudes add up or cancel out when they interfere. This makes it a very useful instrument for calibrating tiny distances. The optical interferometer (that is, one using visible light) really came of age in 1887, when Albert Michelson and Edward Morley exploited its exquisite phase sensitivity to measure the speed of light<sup>2</sup>. Their result — that, however you look at it, light always travels at the same speed — was accurate enough to disprove the existence of the 'luminiferous aether', a universal medium through which light waves were thought to propagate, and thus to pave the way for Albert Einstein's special theory of relativity. The Michelson interferometer still serves today as a test bed for Einstein's general relativity: huge interferometer arms are the basis of observatories dotted around the world looking for signs of gravitational waves, predicted disturbances in the fabric of space-time.

One of these is LIGO, the Laser Interferometer Gravitational-Wave Observatory based at sites in Louisiana and Washington state. It is so sensitive that it can measure a change in the distance travelled by a photon reflected up and down the interferometer's 4-km-long arms, caused by a passing gravitational wave, as small as the diameter of a proton<sup>3</sup>. Even so, over a certain range of frequencies, LIGO is limited by

'shot noise' in the laser beams used to stake out the detector. The origin of this noise is the wave-particle duality of light, which imparts a fluctuation to the light wave's phase ( $\phi$ ) that scales inversely with the square root of the number of photons ( $n$ ) in the laser beam,  $\Delta\phi = 1/\sqrt{n}$ . This uncertainty determines the weakest gravity-wave amplitude that can be detected — the smaller that amplitude is, the better.

The number of photons circulating in the interferometer is proportional to the laser intensity — so, all things being equal, the greater the circulating power, the smaller the phase uncertainty will be. But, in LIGO as in other interferometers, the power cannot be increased indefinitely: mirrors begin to shake from the photon pressure, and the noise starts going up once again. The sweet spot that minimizes noise and maximizes sensitivity gives what is known as the 'standard quantum limit' to the accuracy of an interferometer measurement. For a long time, it was considered the best you could do in experiments involving quantum particles such as photons.

But in 1981, it was pointed out that you could beat the standard limit using quantum trickery<sup>4</sup>. You could simply squeeze the noise in the photon field in such a way that, in the limit of infinite squeezing, you could lower the shot noise and make it depend not on  $1/\sqrt{n}$ , but on  $1/n$  — a quadratic improvement in accuracy at the same light intensity. This lower limit is, in fact, the best you could hope for from the photonic version of the Heisenberg uncertainty principle. This principle is the most fundamental statement of quantum uncertainty, and puts an upper limit on the accuracy to which pairs of variables such as position and momentum (or, equivalently, phase and intensity) can be known. The  $1/n$



**Figure 1 | Towards the Heisenberg limit.** According to quantum theory, the accuracy possible in an interferometric measurement of a phase shift,  $\phi$ , reaches a fundamental 'Heisenberg' limit that scales as the inverse of the number of photons involved in the measurement. But attempts to reach this gold standard experimentally have fallen short. **a**, A promising approach previously investigated<sup>7</sup> sent single Schrödinger's cat 'High-NOON' states consisting of large numbers of photons through the interferometer — all the photons went down one of two arms, but it was impossible to tell which (up cat indicates all photons in the upper path; down cat, all in the lower path).

This approach effectively divided the wavelength of the light (and so its resolving power) by the number of photons in each state, but failed to deliver the hoped-for accuracy, owing to the difficulty of producing and caring for the large cat states. **b**, Higgins and colleagues<sup>1</sup> reach the Heisenberg limit by exploiting a series of smaller, 'kitten' states each involving just one photon. These states are easier to prepare, and are more robust against noise and loss; but the lower flux must be compensated by cycling the states through the interferometer many times and using a complex quantum feedback loop to adjust the phase shift,  $p\phi$ , of the reference beam.

rule is therefore called the Heisenberg limit.

This sounds all very well and good, and people continue to talk about putting squeezed light into LIGO some day. But the fact is that infinite squeezing is hard to come by, and the Heisenberg limit had until now never been realized in practice.

Enter, stage left, the weirdness of quantum entanglement, which occurs when the quantum states of remote particles become intertwined. In 1986, a way was proposed to get close to the Heisenberg limit not with squeezed light, but with quantum-entangled neutrons in a matter-wave interferometer<sup>5</sup>. The entanglement idea percolated along for a number of years, but really gained momentum in the past ten, when people realized that the entanglement approach to interferometry could be implemented using ideas from quantum computing such as error correction and quantum feedback<sup>6</sup>.

A quantum computer is, in essence, a big machine filled with quantum-entangled qubits. A quantum interferometer is also a big machine filled with quantum-entangled particles, and these can be treated as qubits. A popular approach to the phase-estimation problem exploits whacky beasts such as the Schrödinger's cat 'High-NOON' state<sup>7</sup>, in which all the photons are either in one arm of the interferometer or the other, but you can't tell which arm is which (Fig. 1a). In this case, a NOON state of  $n$  photons, each of wavelength  $\lambda$ , acts like a single high-frequency photon of wavelength  $\lambda/n$ . Hence, if one has ten red photons of 500 nm wavelength in an  $n = 10$  NOON state, the result is an entangled red-photon state, but one with the resolving power of an X-ray photon of wavelength 50 nm. The shorter the wavelength, the more accurate the phase

estimation. Much progress was made with such states on both the theoretical and experimental front, and they have got closer to the Heisenberg limit than have squeezed states. But owing to losses in the interferometer and the fragile nature of these states, they have never quite reached the mythical Heisenberg limit<sup>8</sup>.

Until Higgins and colleagues came along<sup>1</sup>. In January 2007, in a theoretical talk at the Physics of Quantum Electronics workshop in Snowbird, Utah, Howard Wiseman from Griffith University in Brisbane, a co-author on the paper, made the remarkable claim that you could get to the ultimate uncertainty limit by sending not Schrödinger's cat through the interferometer, but a bunch of Schrödinger's kittens — single photons. You then compensate for the lower flux and apparent lack of quantum entanglement with an elaborate quantum feedback loop (Fig. 1b). Good luck with that, I remember thinking to myself: applying a feedback loop to single photons at light speed would be technologically impossible any time soon. I am now forced to eat my hat. The authors' optical interferometer, operating at the Heisenberg limit, involves no squeezing, minimal entanglement, and no Schrödinger's cat; the quantum weirdness is in the feedback loop.

This loopy demonstration in fact implements an ingenious phase-estimation algorithm based on quantum computing<sup>9</sup> that uses simple optics to recycle photons through the phase shift to be measured. Although the solution is too low in intensity to be of use in LIGO anytime soon — the largest number of photons the authors used was 378, whereas LIGO has a circulating power of  $10^{14}$  photons per second — the work breaks new ground. It could have other, more immediate applications in areas

such as quantum metrology, quantum imaging and quantum sensing.

So what is the immediate lesson to be learned? That tricks from quantum computing will find their practical near-term implementation in spooky gizmos with scientific and practical importance, but nothing to do with computers at all. Bravo!

Jonathan P. Dowling is at the Hearne Institute for Theoretical Physics, Louisiana State University, Baton Rouge, Louisiana 70803, USA.

e-mail: jdowling@lsu.edu

- Higgins, B. L., Berry, D. W., Bartlett, S. D., Wiseman, H. M. & Pryde, G. J. *Nature* **450**, 393–396 (2007).
- Michelson, A. A. & Morley, E. W. *Phil. Mag.* **24**, 449–463 (1887).
- www.ligo-la.caltech.edu
- Caves, C. M. *Phys. Rev. D* **23**, 1693–1708 (1981).
- Yurke, B. *Phys. Rev. Lett.* **56**, 1515–1517 (1986).
- Lee, H., Kok, P. & Dowling, J. P. *J. Mod. Opt.* **49**, 2325–2338 (2002).
- Bouwmeester, D. *Nature* **429**, 139–141 (2004).
- Nagata, T., Okamoto, R., O'Brien, J. L., Sasaki, K. & Takeuchi, S. *Science* **316**, 726–729 (2007).
- Kitaev, A. Y. *Electr. Coll. Comput. Complex.* **3**, article 3 (1996).

#### Clarification

"Environmental science: Nutrients in synergy" by Eric A. Davidson and Robert W. Howarth (*Nature* **449**, 1000–1001; 2007).

This News & Views article discussed a paper in *Ecology Letters* (doi: 10.1111/j.1461-0248.2007.01113.x; 2007), and included the comment that the results in the paper support the rule-of-thumb that the biological response to phosphorus addition is greater than that to nitrogen addition in freshwater ecosystems. That is true for lake benthos, but not for lake ecosystems as a whole, where the responses to nitrogen and phosphorus are similar.



# Anti-correlation of summer/winter monsoons?

Arising from: G. Yancheva *et al.* *Nature* **445**, 74–77 (2007)

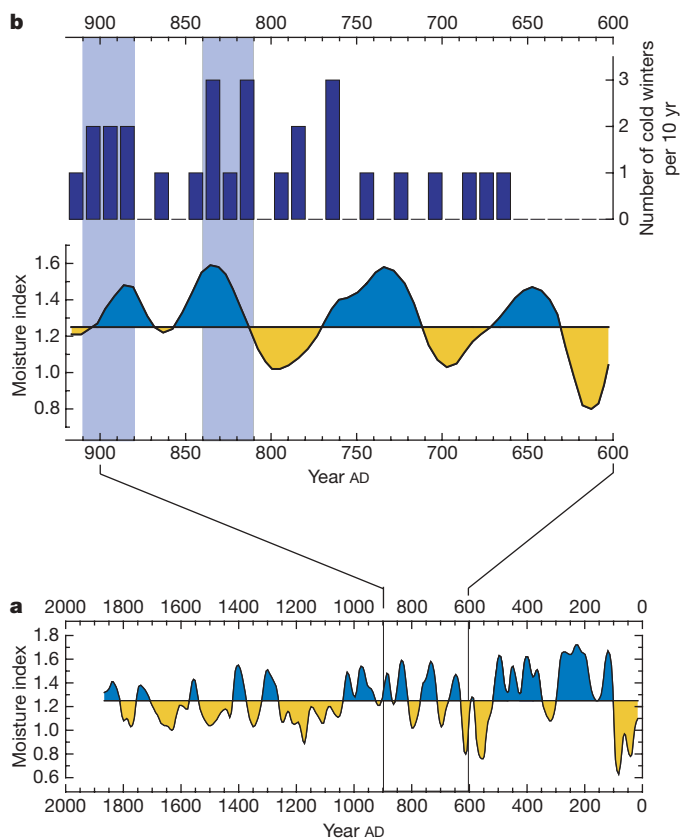
On the basis of the anti-correlation of their palaeoclimatic proxy for the strength of the East Asian winter monsoon from Lake Huguang Maar, China, with stalagmite records of the strength of the summer monsoon, Yancheva *et al.*<sup>1</sup> claim that the strengths of the summer and winter monsoons are anti-correlated on a decadal timescale. They argue that the summer rainfall deficit during AD 700–900 that they infer from their evidence of a stronger winter monsoon, in conjunction with a Tanros battle, led to the collapse of the Tang dynasty (AD 618–907). Using historical climate records, we show here that most cold winters during AD 700–900 were associated with relatively wet summers, indicating that the strengths of the winter and summer monsoons were not negatively correlated during this period.

Yancheva *et al.*<sup>1</sup> deduce an inverse correlation of winter and summer monsoons by comparing centennial- to millennial-scale proxy records of winter monsoon strength from Lake Huguang Maar with records of summer monsoon strength from Chinese caves. Their inference that summer rainfall was low during AD 700–900 contradicts historical climate records for the same period. Chinese historical records can provide records of climatic events on annual to daily timescales<sup>2</sup>. A wetness index for the past 2,000 yr based on 36,750 historical climate records shows that China experienced two wet

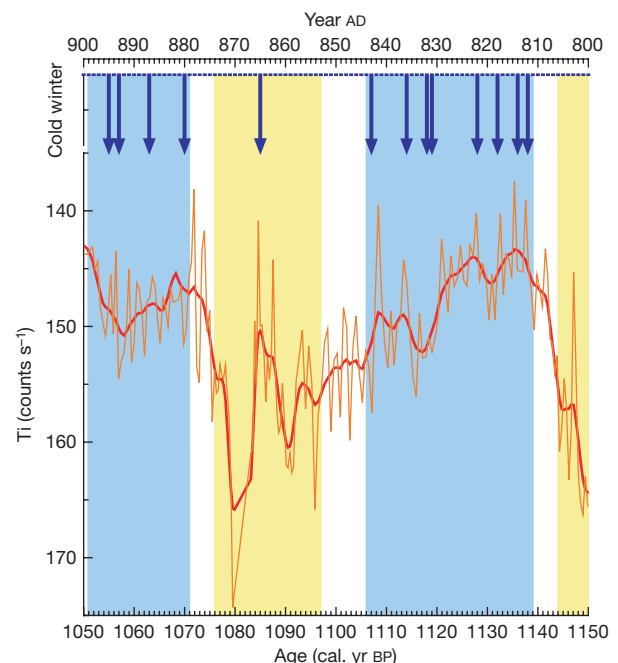
climate phases that bracketed a dry spell during AD 700–900, rather than a “general shift towards drier climate”<sup>2</sup>; wet conditions are recorded in AD 711–770 and 811–1050, with a dry phase in AD 771–810 (ref. 3; Fig. 1), so the last thirty years of the Tang dynasty were relatively wet, not dry.

The historical records show that the climate entered a cold phase from the middle of the eighth century, when the number of cold events between AD 756 and 907 were more than double the number in AD 618–755 (refs 4, 5; Fig. 1), implying that the winter monsoon was strong during the late Tang dynasty. Furthermore, out of 22 cases of severe cold (with anomalously frozen and heavy snow) during AD 700–900, only two occurred in years with a summer rainfall deficit<sup>2</sup>, suggesting that 90% of the harsh winter events were associated with plentiful summer rainfall. This implies that cold winters and wet summers were characteristic of the climatic conditions of the late Tang period, suggesting that the strengths of the winter and summer monsoons were not anti-correlated during this interval.

Yancheva *et al.*<sup>1</sup> interpret their record of titanium (Ti) counts, which has near-annual resolution, as a measure of winter monsoon strength<sup>1</sup>. However, the intervals of increased winter monsoon strength indicated by this record are not always coincident with the intervals indicated by the historical records of severe winter frequency (Fig. 2). During AD 800–900, the highest Ti counts occur around the AD 860s, which is within a period when severe winters were infrequent; according to the historical records, only one



**Figure 1 | Climate fluctuations during the past 2,000 yr based on Chinese historical documentary records.** **a**, The moisture index series at a 5-yr resolution for East China during AD 1–1900; wet and dry intervals are indicated by the blue and yellow shading, respectively (redrawn from ref. 3). **b**, Comparison of the number of severe cold winters per decade (dark blue bars) and the moisture index series for the Tang dynasty (AD 618–907). During the late Tang dynasty (AD 810–907), the increase in the frequency of severe cold winters is associated with plentiful summer rainfall (pale blue shading).



**Figure 2 | Comparison between severe cold winters recorded in Chinese historical documents<sup>2</sup> and the Ti counts from Lake Huguang Maar.** Severe cold winters recorded are shown by blue arrows. The Ti counts from Lake Huguang Maar are shown by the thin red curve (the thick red curve has been smoothed with a 9-point running average, after Yancheva *et al.*<sup>1</sup> from data archived at the World Data Center for Paleoclimatology, Boulder, Colorado, USA) during AD 800–900. The high Ti counts correspond to the lowest frequency of severe cold winters (AD 850–870; yellow shading), while the low Ti counts correspond to increased frequency of cold winters (AD 810–840, 880–900; blue shading). The winter monsoon strength inferred from Ti counts contradicts that based on the occurrence of cold winters in ancient Chinese records.

(AD 865) occurred between AD 850 and 880. In contrast, AD 810–840 witnessed seven very severe winters, including the sea ice event of AD 822 (ref. 2), which is often considered to be the coldest interval during the Tang dynasty. According to the interpretation of Yancheva *et al.*<sup>1</sup>, the lower Ti counts recorded in Lake Huguang Maar<sup>1</sup> during this interval indicate that the winter monsoon was weak during this interval (Fig. 2), in conflict with the historical evidence. Their proxy-based assessment of intervals of increased winter monsoon strength during this period should therefore be called into question.

Furthermore, the intervals of increased winter monsoon strength in AD 810–840 and AD 880–900, indicated by an increase in the frequency of cold winters recorded in the historical records, were both associated with plentiful summer rainfall, whereas the interval of reduced winter monsoon strength between about AD 850 and 870 was associated with a summer rainfall deficit (Fig. 1). This casts further doubt on the proposal of Yancheva *et al.* that the strengths of the winter and summer monsoons were anti-correlated at a multi-decadal scale during AD 600–900.

Yancheva *et al.* say that the Tang dynasty began to ebb in the eighth century, after a defeat by the Arab army in AD 751, but that border battle was not a major event; it is better known for spreading ancient Chinese paper-making techniques to the Arabic countries by prisoners<sup>6</sup>. Instead, most Chinese historians agree that the decline of Tang

dynasty began with the An-Shi riot in AD 755, owing to the resulting disorders, which exhausted resources<sup>6</sup>. The view of Yancheva *et al.* on the historical significance of the border conflict in AD 751 is therefore in doubt.

**De'er Zhang<sup>1</sup> & Longhua Lu<sup>2</sup>**

<sup>1</sup>National Climate Centre, China Meteorological Administration, Beijing 100081, China.

e-mail: derzhang@cma.gov.cn

<sup>2</sup>Chinese Academy of Meteorological Sciences, Beijing 100081, China.

**Received 17 February; accepted 24 August 2007.**

1. Yancheva, G. *et al.* Influence of the intertropical convergence zone on the East Asian monsoon. *Nature* **445**, 74–77 (2007).
2. Zhang, D. E. A *Compendium of Chinese Meteorological Records of the Past 3,000 Years* [in Chinese with English summary] 327–371 (Jiangsu Education Press, Nanjing, 2004).
3. Zheng, S. Z., Zhang, F. C. & Gong, G. F. in *Collection of Papers Regarding Climate Change and Extra Long-Term Prediction* [in Chinese] 29–32 (compiled by China Meteorological Administration, Science Press, Beijing, 1977).
4. Man, Z. M. Climate in the Tang dynasty of China: Discussion of its evidence [in Chinese with English abstract]. *Quat. Sci.* **22**, 20–30 (1998).
5. Zhang, D. E. Environmental change and agricultural development kept in historical documentary records for Northwest China. *Adv. Clim. Change Res.* **2** (suppl. 1), 28–34 (2006).
6. Fan, W. L. *General History of China* [in Chinese] Vol. III, 113–427; 376–377 (People's Publishing House, Beijing, 1978).

doi:10.1038/nature06338

## Yancheva *et al.* reply

**Replying to: De'er Zhang & Longhua Lu *Nature* 450, doi:10.1038/nature06338 (2006)**

Zhang and Lu<sup>1</sup> argue that Chinese historical climate records contradict certain of our interpretations<sup>2</sup> based on Lake Huguang Maar sediment records. Interpreting these records as an indicator for winter monsoon winds and Chinese cave records as an indicator for summer monsoon rainfall<sup>3</sup>, we observed<sup>2</sup> an inverse relationship between winter and summer monsoons on a millennial timescale over the past 16,000 yr. In sediments deposited during the period of Classical Chinese dynastic history, we found evidence for a temporal coincidence between winter monsoon strengthening and the terminations of important dynasties<sup>2</sup>. Extrapolating the inverse monsoon relationship to these multidecadal timescales, we suggested that reduced summer rainfall contributed to dynastic terminations, including that of the Tang (AD 618–907). Zhang and Lu<sup>1</sup> challenge the validity of the summer/winter monsoon relationship on the grounds that historical records indicate that relatively cold winters tended to be associated with relatively wet summers over the period they considered. They argue that the Tang dynasty decline was associated with cold winters, in agreement with our findings<sup>2</sup>, but they find no evidence for rainfall changes having contributed to the decline of the Tang.

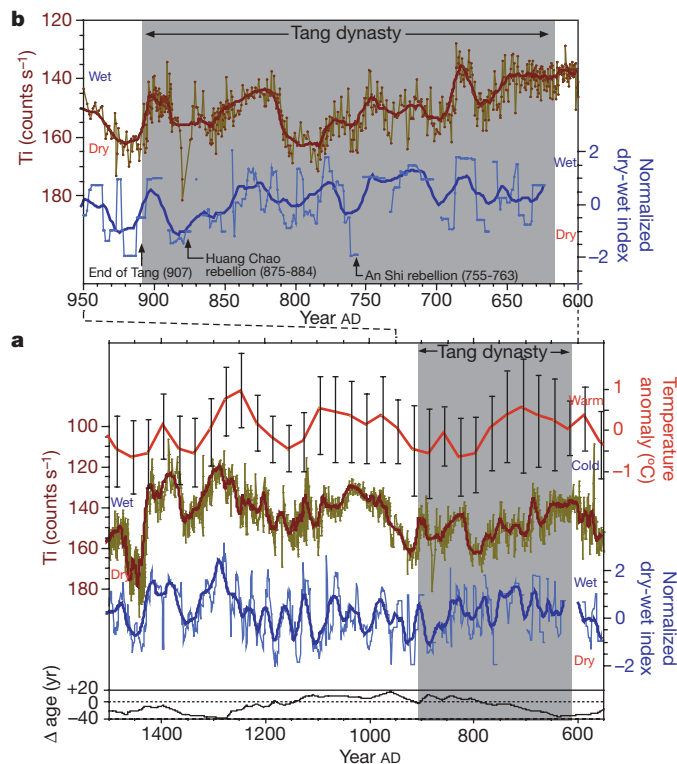
We compare our Lake Huguang Maar record during AD 550–1550 with two historical reconstructions (Fig. 1): a hydrologic reconstruction from the Jiang-Nan area in southeast China (25–31° N)<sup>4</sup>, the most proximal record of which we are aware, and a lower-resolution, winter, half-year, temperature reconstruction for eastern China<sup>5</sup>. To clarify similarities between the Lake Huguang Maar sediment record<sup>2</sup> and the historical climate records<sup>4,5</sup>, we have allowed for stratigraphic adjustments in the former by no more than 40 yr (Fig. 1). The higher-resolution historical reconstruction of wet/dry conditions in south-east China<sup>4</sup> corresponds well with weak/strong winter monsoon

intervals, as indicated by our lake records and by the historical winter temperature record. Its correspondence with the wet/dry record of Zhang and Lu is uncertain; the apparent differences between the two wet/dry historical reconstructions may derive from geographic variability or error in the reconstructions. The previously observed longer-term anti-correlation between winter and summer monsoons seems to apply to a significant portion of the (multi)decadal-scale variability investigated here.

We acknowledge that much of the last 20 yr of the late Tang dynasty was characterized by wetter conditions<sup>1</sup>. Nevertheless, the end of the Tang dynasty at AD 907, which occurs at a sharp increase in Ti content at Lake Huguang Maar<sup>2</sup>, is also marked by a sharp drying in the unsmoothed historical reconstruction of Zheng *et al.*<sup>4</sup> (Fig. 1, thin blue lines), as well as in the record presented by Zhang and Lu<sup>1</sup>. Thus, the historical data suggest that the Tang dynasty collapsed during a period of both cold winters and low and abruptly dropping rainfall (Fig. 1). This may be critical: the rate of climate change could be as important as mean climate from the point of view of social and political impact, as rapid change might disconnect social policies from environmental realities. We acknowledge the clarification of Zhang and Lu that the An-Shi rebellion was more important than the AD 751 Tanros battle in the weakening of the Tang empire.

Both temperature and rainfall seem to have had an important influence on Classical Chinese dynastic changes, including the decline of the Tang dynasty<sup>1,4,5</sup>. Although the reconstructions are somewhat ambiguous on the point, the inverse correlation between winter and summer monsoons observed on millennial timescales<sup>2</sup> does describe some of the decadal and multi-decadal climate variability in Chinese dynastic history.





**Figure 1 | Comparison of Chinese historical climate records<sup>4,5</sup> with the Lake Huguang Maar sediment record<sup>2</sup>.** Comparisons are shown for the time intervals AD 550–1500 (**a**) and AD 600–950 (**b**). In **a**, the red line shows the reconstruction of winter half-year temperatures in eastern China<sup>5</sup> (error bars show uncertainties associated with historical data); **a** and **b** show the Lake Huguang Maar titanium record<sup>2</sup> (brown: original data and 15-point running mean), and the reconstruction of hydrologic conditions from the Jiang-Nan area in southeast China<sup>4</sup> (blue: original data and 15-point running mean). The latter are expressed as a normalized dry-wet index; data are connected by a thin blue line only where data points are available from consecutive years. High Ti values (plotted downward) are interpreted to reflect stronger winter monsoon winds<sup>2</sup>. We have allowed for stratigraphic adjustments in the Huguang Maar sediment record<sup>2</sup> that are consistent within error with the published age model. None of these were more than 40 yr, and most were much less, including during the interval of the Tang dynasty (black curve at bottom of **a** shows the adjusted timescale minus the <sup>14</sup>C-based timescale as described in Yancheva *et al.*<sup>2</sup>, such that the vertical scale matches that of the x-axis).

Gergana Yancheva<sup>1</sup>, Norbert R. Nowaczyk<sup>1</sup>, Jens Mingram<sup>1</sup>, Peter Dulski<sup>1</sup>, Georg Schettler<sup>1</sup>, Jörg F. W. Negendank<sup>1</sup>, Jiaqi Liu<sup>2</sup>, Daniel M. Sigman<sup>3</sup>, Larry C. Peterson<sup>4</sup> & Gerald H. Haug<sup>5</sup>

<sup>1</sup>GeoForschungsZentrum Potsdam, Telegrafenberg, D-14473 Potsdam, Germany.

<sup>2</sup>Institute of Geology and Geophysics, Chinese Academy of Sciences, PO Box 9825, Beijing 100029, China.

<sup>3</sup>Department of Geosciences, Princeton University, Princeton, New Jersey 08544, USA.

<sup>4</sup>Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, Florida 33149, USA.

<sup>5</sup>Geological Institute, Department of Earth Sciences, ETH Zürich, CH-8092 Zürich, Switzerland.

e-mail: gerald.haug@erdw.ethz.ch

1. Zhang, D. E. & Lu, L. H. Anti-correlation of summer/winter monsoons? *Nature* **450**, doi:10.1038/nature06338 (2007).
2. Yancheva, G. *et al.* Influence of the intertropical convergence zone on the East Asian monsoon. *Nature* **455**, 74–77 (2007).
3. Dykoski, C. A. A high-resolution, absolute-dated Holocene and deglacial Asian monsoon record from Dongge Cave, China. *Earth Planet. Sci. Lett.* **233**, 71–86 (2005).
4. Zheng, J. *et al.* Precipitation variability and extreme events in eastern China during the past 1500 years. *Terr. Atmos. Ocean. Sci.* **17**, 579–592 (2006).
5. Ge, Q. *et al.* Winter half-year temperature reconstruction for the middle and lower reaches of the Yellow River and the Yangtze River, China, during the past 2000 years. *Holocene* **13**, 933–940 (2003).

doi:10.1038/nature06339

# Record of winter monsoon strength

Arising from: G. Yancheva *et al.* *Nature* **455**, 74–77 (2007).

The Asian summer monsoon has been precisely reconstructed from the high-resolution record from the speleothem<sup>1</sup>, but reconstruction of the Asian winter monsoon is less satisfactory. Yancheva *et al.*<sup>2</sup> provide such a reconstruction for the last 16,000 years from the titanium (Ti) content of the sediments of Lake Huguang Maar in coastal South China. However, we argue that the Ti is likely to have come mainly from the catchment and so the Ti content may instead be related to the hydrology of the lake.

Considering the small catchment and densely vegetated inner slope of the lake, Yancheva *et al.*<sup>2</sup> overlook the Ti from the catchment through runoff. However, weathering and erosion in coastal South China are intensive, owing to high temperatures and heavy precipitation. The steep slopes surrounding the lake<sup>3</sup> would promote surface erosion further. Sandy sediments in the lake<sup>4</sup> are strong evidence for significant terrigenous input through runoff. This is strengthened by records from Tianyang Maar in the Leizhou Peninsula<sup>5</sup> and Shuangchi Maar in the north of Hainan<sup>6</sup>, which indicate that terrigenous input through runoff may be common for Maars in coastal South China. The basalts and laterite in the Leizhou peninsula, which are dominant in the catchment, have a high Ti content<sup>7</sup> (Table 1) and so should be important sources of Ti for the lake.

More importantly, the flux of the lithogenic materials to the lake, 5–10 and 10–20 mg cm<sup>-2</sup> yr<sup>-1</sup> during the Holocene and glacial times, respectively<sup>2</sup>, is too high to be explained by a wind-blown mechanism. Although it is comparable with an estimated modern dust flux of 4.6 mg cm<sup>-2</sup> yr<sup>-1</sup> in coastal South China<sup>2</sup>, the dust flux was probably overestimated, given the atmospheric dust concentration in coastal South China<sup>8</sup> and its settling velocity<sup>9</sup>. In addition, modern observation indicates that mineral dust flux on the Loess plateau ranges between 15 and 36 mg cm<sup>-2</sup> yr<sup>-1</sup> at sites close to dust sources and between 9 and 10 mg cm<sup>-2</sup> yr<sup>-1</sup> at other sites, but only about 25% of the particles are smaller than 6 µm (ref. 10). Particles larger than 6 µm generally have short atmospheric lifetimes owing to gravitational settling<sup>11</sup>. Lake Huguang Maar is about 1,500 km south of the Loess plateau and is not on a main dust trajectory. It should have a dust deposition rate much smaller than 2.5 mg cm<sup>-2</sup> yr<sup>-1</sup>. This is consistent with model-derived deposition rates of <0.2 to 0.5 mg cm<sup>-2</sup> yr<sup>-1</sup> at this site<sup>11,12</sup>. This evidence disproves wind-blown dust from remote North China as the main source of Ti to Lake Huguang Maar.

The Ti content of local basalts and laterite is apparently higher than in the loess deposit in the Loess plateau (Table 1)<sup>7,13</sup>. A source of Ti from the catchment is therefore more reasonable. This explanation is supported by the correspondence of high Ti content with the occurrence of sandy sediments during two cold, dry phases, the Younger Dryas episode and the period before the Bølling–Allerød warming<sup>2,4</sup>.

**Table 1 | Ti content in basalt and laterite and in loess-palaeosol**

	Laterite range (%)	Laterite average (%)	Basalt (%)
Site 9	1.67–1.76	1.71	1.12
Site 10	1.53–1.61	1.57	1.05
Site 11	1.24–1.57	1.47	
Site 12	0.93–1.13	1.03	0.88
Site 13	1.16–1.51	1.26	
	Range (%)	Average (%)	
Loess	0.43–0.53	0.48	
Palaeosol	0.42–0.51	0.44	

The Ti contents are calculated according to the TiO<sub>2</sub> contents in the basalt and laterite on the Leizhou Peninsula, coastal South China<sup>7</sup>, and in the loess-palaeosol deposition at Luochuan, North China<sup>13</sup>.

It would also account for some puzzles arising from the connection between Ti content and the Asian winter monsoon. For example, although the Ti content indicates a southward movement of the intertropical convergence zone 7.8 kyr ago<sup>2</sup>, some places further north experienced a stronger summer monsoon until about 6 kyr ago<sup>14,15</sup>; the Ti content during the Holocene is comparable with those during the Younger Dryas and the pre-Bølling–Allerød periods<sup>2</sup>, but other records indicate that the summer monsoon was much stronger and that the winter monsoon was much weaker during the Holocene than during the two cold, dry periods<sup>15,16</sup>.

The correspondence between the higher Ti content and occurrence of coarser sediments during the Younger Dryas and the pre-Bølling–Allerød periods<sup>2,4</sup> indicates a connection between the Ti content and the hydrology of the lake. The higher Ti content during the period 5.2–7.8 kyr ago<sup>2</sup> may have been caused by a similar mechanism because precipitation may have been lower for this period<sup>3</sup>. Precipitation in coastal South China during 7.2–2.7 kyr ago may also have been reduced in Shuangchi Maar<sup>6</sup>. If Ti was transported into the lake mainly in particle and colloidal phases, increased input through runoff and reduced water depth could have contributed to the higher Ti content recorded in these cores<sup>2</sup>.

Our alternative interpretation is therefore that the Ti content may have been controlled by the hydrology of the lake, rather than by the Asian winter monsoon. Whether, and how, this mechanism could influence proxies such as total organic carbon and magnetic properties, however, needs further investigation.

Houyun Zhou<sup>1,2</sup>, Huazheng Guan<sup>1,3</sup> & Baoquan Chi<sup>1,3</sup>

<sup>1</sup>Guangzhou Institute of Geochemistry, Chinese Academy of Sciences, Guangzhou 510640, China.

<sup>2</sup>The Earth Dynamic System Research Center, National Cheng-Kung University, Tainan 701, Taiwan.

e-mail: hyzhou@gig.ac.cn

<sup>3</sup>The Graduate School, Chinese Academy of Sciences, Beijing 100039, China.

Received 25 January; accepted 20 September 2007.

- Wang, Y.-J. *et al.* A high-resolution absolute-dated late Pleistocene monsoon record from Hulu Cave, China. *Science* **294**, 2345–2348 (2001).
- Yancheva, G. *et al.* Influence of the intertropical convergence zone on the East Asian monsoon. *Nature* **455**, 74–77 (2007).
- Mingram, J. *et al.* The Huguang maar lake—a high-resolution record of palaeoenvironmental and palaeoclimatic changes over the last 78,000 years from South China. *Quat. Int.* **122**, 85–107 (2004).
- Wang, W.-Y. *et al.* The two-step monsoon changes of the last deglaciation recorded in tropical Maar Lake Huguangyan, southern China. *Chin. Sci. Bull.* **45**, 1529–1532 (2000).
- Zheng, Z. & Lei, Z.-Q. A 400,000 year record of vegetational and climatic changes from a volcanic basin, Leizhou Peninsula, southern China. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **145**, 339–362 (1999).
- Zheng, Z. *et al.* High-resolution records of Holocene from the Shuangchi Maar Lake in Hainan Island. *Chin. Sci. Bull.* **48**, 497–502 (2003).
- Xu, Y.-F., Zhu, Z.-Y., Wen, G.-G., Liang, J.-P. & Luo, S.-W. Geochemistry and soil-forming environment of the red soil section in the southern Leizhou Peninsula, Guangdong Province. *Geochimica* **29**, 402–408 (2000).
- Cao, J.-J. *et al.* Characterization of dust storms to Hong Kong in April 1998. *Water Air Soil Pollut. Focus* **3**, 213–229 (2003).
- Reid, J. S. *et al.* Analysis of measurements of Saharan dust by airborne and ground-based remote sensing methods during the Puerto Rico Dust Experiment (PRIDE). *J. Geophys. Res.* **108** (D19), 8586, doi:10.1029/2002JD002493 (2003).
- Sun, D.-H., Chen, F.-H., Bloemendal, J. & Su, R.-X. Seasonal variability of modern dust over the Loess Plateau of China. *J. Geophys. Res.* **108** (D21), 4665, doi:10.1029/2003JD003382 (2003).
- Ginoux, P. *et al.* Sources and distributions of dust aerosols simulated with the GOCART model. *J. Geophys. Res.* **106** (D17), 20255–20273 (2001).
- Mahowald, N. *et al.* Dust sources and deposition during the last glacial maximum and current climate: A comparison of model results with paleodata



- from ice cores and marine sediments. *J. Geophys. Res.* **104** (D13), 15895–15916 (1999).
13. Gallet, S. & Jahn, B.-M. Geochemical characterization of the Luochuan loess-paleosol sequence, China, and paleoclimatic implications. *Chem. Geol.* **133**, 67–88 (1996).
  14. Shao, X.-H. et al. Long-term trend and abrupt events of the Holocene Asian monsoon inferred from a stalagmite  $\delta^{18}\text{O}$  record from Shennongjia in Central China. *Chin. Sci. Bull.* **51**, 221–228 (2006).

15. Dykoski, C. A. et al. A high-resolution, absolute-dated Holocene and deglacial Asian monsoon record from Dongge Cave, China. *Earth Planet. Sci. Lett.* **233**, 71–86 (2005).
16. Ren, J.-Z., Ding, Z.-L., Liu, D.-S., Sun, J.-M. & Zhou, X.-Q. Climatic changes on millennial time scales - Evidence from a high-resolution loess record. *Sci. Chin. Ser. D* **39**, 449–459 (1996).

doi:10.1038/nature06408

## Yancheva et al. reply

Replying to: H. Zhou et al. *Nature* **450**, doi: 10.1038/nature06408 (2007).

Zhou et al.<sup>1</sup> raise the possibility that the titanium (Ti) record at Lake Huguang Maar is controlled by local erosion and runoff to the lake, or through hydrological changes in the lake such as level fluctuations, rather than by changes in the inputs of airborne material<sup>2</sup>. The authors come to this conclusion by considering the Ti record in isolation. They ignore the redox-sensitive parameters of the S-ratio, total organic-matter content, the Mn/Fe ratio, and magnetic susceptibility records. We believe that their case against our interpretation of the Ti record is weak, and that their interpretation can be ruled out if the other measurements are taken into account<sup>2</sup>.

Zhou et al.<sup>1</sup> argue that the high Ti content during times of weak summer monsoon, such as the Younger Dryas cold period, may be the result of runoff from the catchment. On the basis of runoff alone, one would expect the opposite signal, with lower Ti values in times of reduced rainfall and runoff, and not higher values. The authors argue that lake-level changes may explain the higher Ti values during drier climates. However, major lake-level changes can be excluded in the case of Lake Huguang Maar for the study interval.

The lake is today about 21 m deep, and it is characterized by a strong thermocline at about 10 m water depth. Below the thermocline, the lake is suboxic. Our transect of cores spans the depth range between 10 and 20 m water depth<sup>2</sup>. For all cores at all depths, we observe a continuous sediment sequence during the past 20,000 years, as indicated by parallel high-resolution magnetic susceptibility records in all cores. This argues that the lake level has not fluctuated by as much as 10 m, because this would have exposed the shallowest core to the atmosphere. Moreover, suboxic to anoxic conditions are recorded in our magnetic susceptibility and S-ratio records throughout the record at all cores and at all water depths. Even a modest drop in lake level would have exposed the shallowest core (at 10 m water depth) to the oxic lake conditions in the epilimnion above the thermocline, which would have markedly altered the redox sensitive signals of this core.

Zhou et al.<sup>1</sup> argue that the correspondence of higher Ti with coarser sediments during the Younger Dryas and the pre-Bølling/Allerød periods indicates a connection between Ti content and the hydrology of the lake. However, microscopic quantification of minerogenic matter of local origin shows only scattered silt-sized

minerogenic particles, as well as more prominent siderite enrichments, but without any major changes, even when including the glacial termination in the analysis. There are no sandy sediments in the entire Lake Huguang Maar section: the interpretation of Wang et al.<sup>3</sup> is incorrect.

One aspect of the comment is valid and important. Given the current understanding of large scale dust distributions<sup>4</sup>, regional dust deposition seems to be significantly lower than siliciclastic flux rates in Lake Huguang Maar sediments. This discrepancy may be explained by sediment focusing in similar lake environments, which has been measured as up to fivefold<sup>5</sup>.

**Gergana Yancheva<sup>1</sup>, Norbert R. Nowaczyk<sup>1</sup>, Jens Mingram<sup>1</sup>, Peter Dulski<sup>1</sup>, Georg Schettler<sup>1</sup>, Jörg F. W. Negendank<sup>1</sup>, Jiaqi Liu<sup>2</sup>, Daniel M. Sigman<sup>3</sup>, Larry C. Peterson<sup>4</sup> & Gerald H. Haug<sup>5</sup>**

<sup>1</sup>GeoForschungsZentrum Potsdam, Telegrafenberg, D-14473 Potsdam, Germany.

<sup>2</sup>Institute of Geology and Geophysics, Chinese Academy of Sciences, PO Box 9825, Beijing 100029, China.

<sup>3</sup>Department of Geosciences, Princeton University, Princeton, New Jersey 08544, USA.

<sup>4</sup>Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, Florida 33149, USA.

<sup>5</sup>Geological Institute, Department of Earth Sciences, ETH Zürich, CH-8092 Zürich, Switzerland.

e-mail: gerald.haug@erdw.ethz.ch

1. Zhou, H., Guan, H. & Chi, B. Records of winter monsoon strength. *Nature* **450**, doi: 10.1038/nature06408 (2007).
2. Yancheva, G. et al. Influence of the intertropical convergence zone on the East Asian monsoon. *Nature* **455**, 74–77 (2007).
3. Wang, W.-Y. et al. The two-step monsoon changes of the last deglaciation recorded in tropical Maar Lake Huguangyan, southern China. *Chin. Sci. Bull.* **45**, 1529–1532 (2000).
4. Mahowald, N. et al. Dust sources and deposition during the last glacial maximum and current climate: A comparison of model and results with paleodata from ice cores and marine sediments. *J. Geophys. Res.* **104** (D13), 15895–15916 (1999).
5. Schettler, G., Liu, Q., Mingram, J. & Negendank, J. F. W. Palaeovariations in the East-Asian monsoon regime geochemically recorded in varved sediments of lake Sihailongwan (northeast China, Jilin province). Part 1: Hydrological conditions and dust flux. *J. Paleolimnol.* **35**, 239–270 (2006).

doi:10.1038/nature06409

# Legionella pneumophila proteins that regulate Rab1 membrane cycling

Alyssa Ingmundson<sup>1</sup>, Anna Delprato<sup>2</sup>, David G. Lambright<sup>2</sup> & Craig R. Roy<sup>1</sup>

**Rab1 is a GTPase that regulates the transport of endoplasmic-reticulum-derived vesicles in eukaryotic cells. The intracellular pathogen *Legionella pneumophila* subverts Rab1 function to create a vacuole that supports bacterial replication by a mechanism that is not well understood. Here we describe *L. pneumophila* proteins that control Rab1 activity directly. We show that a region in the DrrA (defect in Rab1 recruitment A) protein required for recruitment of Rab1 to membranes functions as a guanine nucleotide dissociation inhibitor displacement factor. A second region of the DrrA protein stimulated Rab1 activation by functioning as a guanine nucleotide exchange factor. The LepB protein was found to inactivate Rab1 by stimulating GTP hydrolysis, indicating that LepB has GTPase-activating protein activity that regulates removal of Rab proteins from membranes. Thus, *L. pneumophila* encodes proteins that regulate three distinct biochemical reactions critical for Rab GTPase membrane cycling to redirect Rab1 to the pathogen-occupied vacuole and to control Rab1 function.**

Rab proteins are molecular switches that have an important role in organizing membranes to facilitate transport and fusion of vesicular carriers<sup>1</sup>. Rab GTPases localize and control membrane dynamics on specific subcellular organelles. Successful transport and maturation of vesicles coincides with a change in the composition of Rab proteins—a process called Rab conversion<sup>2</sup>. Several accessory proteins regulate the Rab GTPase cycle. Prenylated Rab proteins in their inactive GDP-bound state are solubilized from membranes by associating with Rab-guanine nucleotide dissociation inhibitor (Rab GDI)<sup>3–5</sup>. There is evidence that soluble Rab proteins must be released from Rab GDI by a GDI-displacement factor (GDF) before membrane recruitment and activation by a cognate guanine nucleotide exchange factor (GEF)<sup>6</sup>. An active Rab protein recruits specific effectors that have a direct role in promoting membrane transport, vesicle tethering and fusion<sup>7,8</sup>. To complete the process of membrane cycling, Rab proteins are inactivated by cognate GTPase-activating proteins (GAPs)<sup>9,10</sup> and are extracted from the membrane by Rab GDI<sup>5</sup>.

After uptake by phagocytic host cells, the bacterial pathogen *L. pneumophila* exploits the activities of the Rab1 protein to direct the maturation of the vacuole in which it resides and to create an endoplasmic-reticulum-derived organelle that supports intracellular replication<sup>11–15</sup>. Because the vacuole, which is initially derived from plasma membrane, is remodelled into an endoplasmic-reticulum-like compartment, host proteins involved in regulating endoplasmic reticulum to Golgi transport are recruited to the *L. pneumophila*-containing vacuole (LCV) through the activities of *L. pneumophila* proteins delivered into the host cytosol by a bacterial type IV secretion apparatus called the Dot/Icm system<sup>14–18</sup>. Rab1 is one of the host regulatory factors recruited to the LCV<sup>14,15</sup>. Although the mechanism by which Rab1 is recruited to the LCV is not well described, it was determined that the association of Rab1 with the LCV requires the *L. pneumophila* protein DrrA (also known as SidM)<sup>19,20</sup>.

Previous biochemical studies have shown that DrrA functions as a Rab1-specific GEF<sup>19,20</sup>. Nucleotide exchange reactions using purified Rab1 protein revealed that a carboxy-terminal region of DrrA consisting of amino acid residues 451–647, DrrA(451–647), had the same

GEF activity and specificity for Rab1 as the full-length DrrA protein<sup>19</sup>. Importantly, a DrrA derivative, consisting of residues 61–647, DrrA(61–647), could recruit Rab1 to the plasma membrane when expressed ectopically in mammalian cells, whereas the plasma-membrane-localized DrrA(451–647) protein could not recruit Rab1 (ref. 19). These data indicated that deletion of residues 61–450 in DrrA eliminated a second biochemical activity essential for the membrane recruitment and activation of Rab1 from the host cytosol. Here we show that region 61–450 in DrrA is necessary to displace Rab GDI from a complex containing Rab1, indicating that DrrA has GDF activity. In addition, we show that the *L. pneumophila* LepB protein is a Rab1 GAP capable of inactivating Rab1.

## DrrA has both GDF and GEF activity

The observation that the GEF domain alone is not sufficient for DrrA-mediated recruitment of Rab1 to the plasma membrane<sup>19</sup> indicated that DrrA might also function as a GDF. We reasoned that, if DrrA is able to displace Rab GDI *in vivo*, ectopic production of DrrA in eukaryotic cells should decrease the proportion of Rab1 that is bound to Rab GDI. Rab1 tagged with the Flag epitope (3×Flag–Rab1) was affinity-purified using anti-Flag agarose from HEK293 cells producing Rab GDI and green fluorescent protein (GFP)-tagged DrrA proteins. Rab GDI was not detected in association with 3×Flag–Rab1 isolated from cells producing GFP–DrrA(61–647) (Fig. 1a), a derivative shown previously to recruit Rab1 to the plasma membrane<sup>19</sup>. In contrast, Rab GDI was associated with 3×Flag–Rab1 isolated from cells producing either GFP alone or GFP–DrrA(451–647) with an intact GEF domain (Fig. 1a). These data indicate that the DrrA protein can interfere with Rab1–Rab-GDI complex formation by a mechanism that requires residues outside the GEF domain.

To determine whether DrrA could displace Rab GDI *in vitro*, purified DrrA was added to 3×Flag–Rab1–Rab-GDI complex attached to anti-Flag agarose, and Rab GDI displacement into the supernatant was measured over time. Rab GDI was efficiently displaced when DrrA protein was added to the reaction (Fig. 1b, c).

<sup>1</sup>Section of Microbial Pathogenesis, Yale University School of Medicine, Boyer Center for Molecular Medicine, 295 Congress Avenue, New Haven, Connecticut 06536, USA. <sup>2</sup>Program in Molecular Medicine and Department of Biochemistry and Molecular Pharmacology, UMASS Medical School Two Biotech, 373 Plantation Street, Worcester, Massachusetts 01605, USA.



DrrA also displaced Rab GDI from a complex containing 3×Flag–Rab2 (Supplementary Information and Supplementary Fig. 1), indicating that DrrA may function in a similar manner to the eukaryotic protein Rab acceptor 1 (prenylated) (Rabac1, also known as Yip3 in yeast and PRA1 in human) and promote the displacement of multiple Rab proteins from the complex with Rab GDI<sup>21</sup>. DrrA(451–647) did not stimulate Rab GDI release, consistent with Rab GDI displacement activity requiring residues outside of the GEF domain. Importantly, DrrA(1–500), which lacks Rab1 GEF activity<sup>19</sup>, was sufficient to displace Rab GDI from Rab1, indicating that Rab1 nucleotide exchange is not required for Rab GDI displacement by DrrA. Thus, Rab GDI displacement and Rab1 guanine nucleotide exchange are separate activities mediated by the DrrA protein.

Interactions between Rab GDI and a Rab protein can interfere with GEF activation because Rab GDI binding directly engages the Rab switch regions and inhibits GDP release<sup>22,23</sup>. A GDF is predicted to recognize both Rab GDI and the Rab protein, and release Rab GDI from the complex, to enable a GEF to access the switch regions and catalyse nucleotide exchange. Our data indicated that DrrA possesses both GDF and GEF activity, predicting that DrrA should be able to activate Rab1 even when Rab1–Rab-GDI is used as a substrate. The sensitized emission accompanying tryptophan to 2'(3')-O-(N-methylanthraniloyl) (mant) fluorescence resonance energy transfer was used to measure the exchange of GDP for mant-GTP using Rab1–Rab-GDI as a substrate. Full-length DrrA activated Rab1, whereas the DrrA(451–647) protein containing the GEF domain but deficient in GDF activity was unable to activate Rab1 when the Rab1–Rab-GDI complex was used as a substrate (Fig. 1d). These data are in contrast to previous data showing that full-length DrrA and DrrA(451–647) had similar catalytic activities in exchange reactions for Rab1 alone, and demonstrate that when Rab1–Rab-GDI is used as a substrate, Rab GDI displacement is essential for DrrA-mediated activation of Rab1.

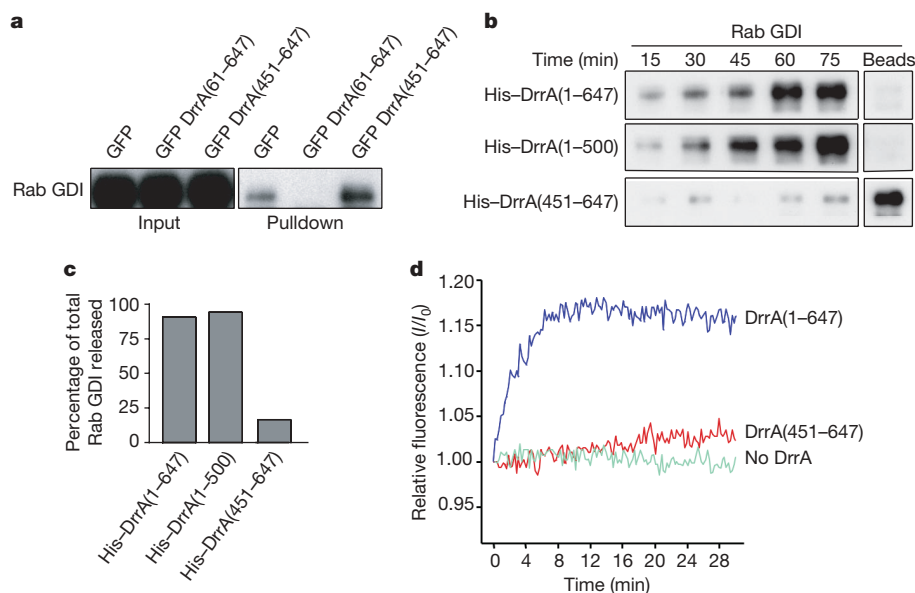
### LepB binds Rab1–GTP

Because Rab1 on the LCV would be available for binding to other *L. pneumophila* proteins after DrrA activation, we investigated

whether any Dot/Icm-translocated proteins of unknown function could interact with Rab1 in a GTP-bound conformation. This was done using the GTP-locked glutathione S-transferase (GST)–Rab1(Q70L) protein (in which Gln 70 was mutated to Leu) to co-precipitate GFP-tagged *L. pneumophila* proteins from transfected mammalian cell lysates. This analysis revealed an interaction between GST–Rab1(Q70L) and the *L. pneumophila* protein LepB (Fig. 2a). The GDP-locked GST–Rab1(S25N) protein did not bind to LepB, indicating that LepB interacts preferentially with active Rab1–GTP (Fig. 2a). LepB did not bind to the GTP-locked GST–Rab7(Q67L), indicating specificity for the interaction between LepB and Rab1. Direct binding between purified His–LepB and GST–Rab1(Q70L) was detected, indicating that this interaction does not require other cellular factors (Fig. 2b).

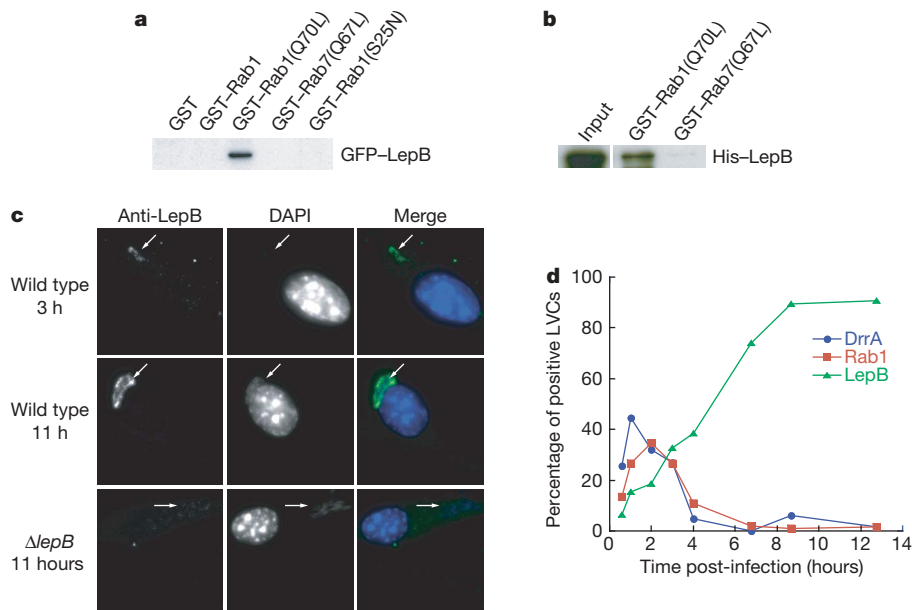
### LepB is translocated shortly after uptake

Previous studies have indicated that LepB is involved in programmed release of *L. pneumophila* from protozoan host cells, and not mammalian cells, by an exocytic process that occurs after bacteria have completed their intracellular replication phase<sup>24</sup>. Thus, the possibility that Rab1 is a target of LepB function was surprising given that Rab1 is conserved in protozoan and metazoan organisms, and because Rab1 is found on the *L. pneumophila* vacuole shortly after bacterial uptake and is removed from the vacuole before bacterial replication initiates<sup>14,15</sup>. Spatial and temporal analysis of LepB localization in *L. pneumophila*-infected macrophages was conducted to determine whether the kinetics of LepB and Rab1 association with the LCV were similar. Immunofluorescence microscopy revealed detectable levels of LepB on LCVs containing a single bacterium at 3 h post infection (Fig. 2c). LepB was also present on mature endoplasmic-reticulum-derived vacuoles containing replicating *L. pneumophila* at 11 h post infection (Fig. 2c). DrrA and Rab1 were present on early LCVs at the 30-min time point after infection, whereas LepB was not detected this early (Fig. 2d). The percentage of LCVs positive for Rab1 peaked at 2 h post infection, a time at which the presence of LepB on the LCVs became apparent. As the frequency of vacuoles staining positive for LepB increased, the percentage of vacuoles staining positive for



**Figure 1 | DrrA has GDF activity that is required for activation of Rab1 bound to Rab GDI.** **a**, The proteins indicated were produced with 3×Flag–Rab1 and Rab GDI in HEK293 cells. The amount of Rab GDI associated with 3×Flag–Rab1 in each lysate (Pull-down) was compared to the total amount in the cell lysate (Input) by immunoblot analysis. **b**, Isolated Rab1–Rab-GDI complex attached to anti-Flag agarose was incubated with the indicated purified His-tagged DrrA proteins. The release of Rab GDI was measured over 75 min after addition of the His–DrrA proteins by

immunoblot analysis of supernatants. Rab GDI that remained associated with the beads at the end of the reaction was eluted in 0.1 M glycine, pH 3.5, and measured by immunoblot analysis (beads). **c**, The amount of Rab GDI released after incubation with the indicated His-tagged DrrA proteins was determined by measuring the intensity of the bands in **b**. **d**, The Rab1–Rab-GDI complex was incubated with the indicated DrrA protein or buffer alone (No DrrA), and increased fluorescence resulting from mant-GTP loading was used to measure Rab1 activation.



**Figure 2 | LepB interacts with GTP-bound Rab1 and is detected on the early LCV.** **a**, The indicated GST fusion proteins were incubated with lysate from HEK293 cells producing GFP-LepB. Binding of GFP-LepB to the precipitated GST fusion protein was measured by immunoblot analysis. **b**, Purified His-tagged LepB was incubated with either GST-Rab1(Q70L) or GST-Rab7(Q67L). Binding of His-LepB to the precipitated GST fusion protein was measured by immunoblot analysis. Input represents 3.7% of the total reaction volume. **c**, Mouse bone-marrow-derived macrophages infected with either wild-type or a  $\Delta lepB$  mutant strain of *L. pneumophila* were fixed either 3 h or 11 h post-infection and stained with a LepB-specific

antibody (green) and 4,6-diamidino-2-phenylindole (DAPI, blue). Arrows in the images obtained by fluorescence microscopy indicate LepB staining of the LCV at 3 h and 11 h for cells infected with wild-type *L. pneumophila*, but not for vacuoles containing the  $\Delta lepB$  mutant. **d**, Mouse bone-marrow-derived macrophages infected with *L. pneumophila* were fixed at the times indicated post infection and stained for DrrA (circles), Rab1 (squares) or LepB (triangles). For the indicated protein at indicated times post infection, at least 100 vacuoles were scored; the numbers represent the percentage of LCVs that stained positive. These results were verified for each protein in two independent experiments.

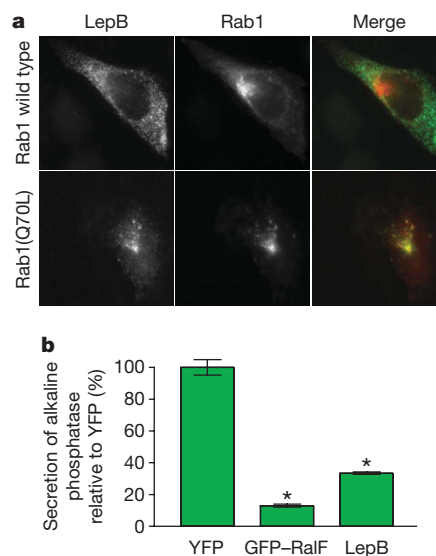
DrrA and Rab1 decreased. At 7 h post infection most of the LCVs stained positive for LepB, but DrrA and Rab1 staining was no longer detectable (Fig. 2d). These data indicate that both Rab1 and LepB are present on the early LCV membrane; however, LepB remains associated with compartments that support *L. pneumophila* replication and Rab1 cycles off.

### LepB disrupts early secretory transport

The finding that LepB is delivered by the Dot/Icm system shortly after bacterial uptake suggests that LepB and Rab1 interact directly on this compartment. To examine *in vivo* co-localization of these proteins in the absence of infection, LepB and Rab1 were produced together in mammalian cells. The LepB protein and GFP-Rab1(Q70L) co-localized in the cell (Fig. 3a); this is consistent with the *in vitro* studies showing that these proteins can interact. In cells producing wild-type Rab1, however, LepB staining was diffuse and showed no appreciable co-localization with red fluorescent protein (RFP)-Rab1 (Fig. 3a), indicating that the GTP-locked derivative of Rab1 forms a more stable complex with LepB than does the wild-type Rab1, which cycles between a GDP- and GTP-bound conformation. To investigate whether LepB might alter Rab1-dependent processes in the eukaryotic cell, the morphology of the Golgi apparatus was examined in cells producing LepB. Overproduction of LepB resulted in the dispersion and fragmentation of the Golgi apparatus, as determined by staining for the matrix protein GM130 (golgi autoantigen, golgin subfamily a, 2, also known as GOLGA2) (Supplementary Fig. 2). Measuring the secretion of an alkaline phosphatase reporter in transfected cells confirmed that production of LepB disrupted the secretory pathway (Fig. 3b).

### LepB is a Rab1 GAP

Effectors and GAPs associate preferentially with Rab proteins in their GTP-bound conformation<sup>7,8</sup>. The appearance of LepB on the LCV at the time at which Rab1 begins to cycle off the LCV, and the



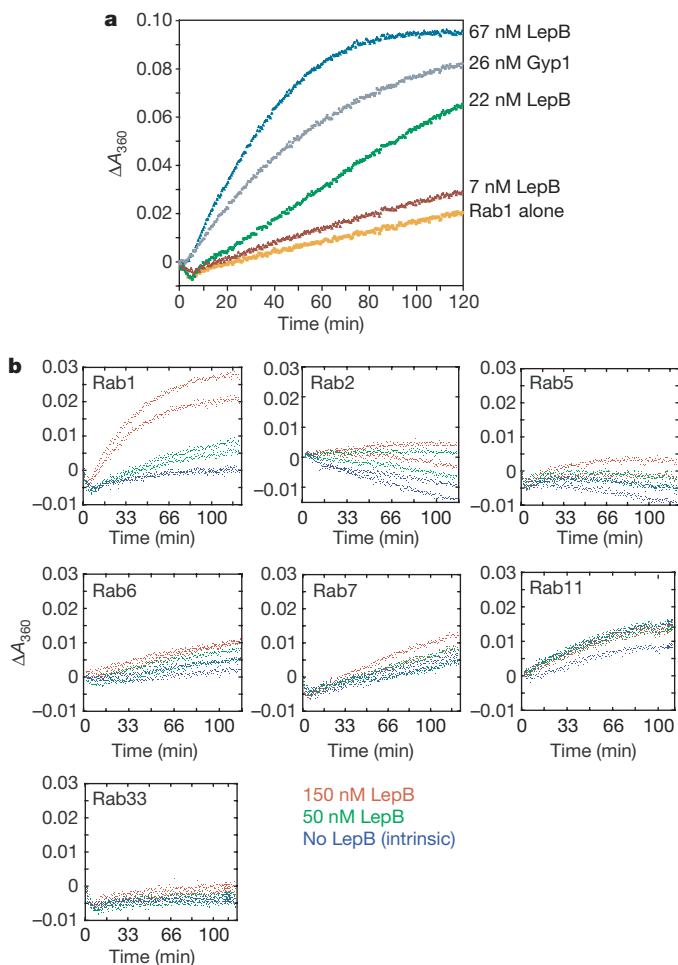
**Figure 3 | LepB interacts with Rab1-GTP and disrupts secretory transport in mammalian cells.** **a**, Ectopically produced LepB protein was localized in CHO Fc $\gamma$ RII cells using a LepB-specific antibody. Fluorescence micrographs show that LepB co-localized with GTP-locked Rab1(Q70L), but did not co-localize with wild-type Rab1 protein. **b**, Secretion of an alkaline phosphatase reporter protein was measured in HEK293 cells producing yellow fluorescent protein (YFP), the *L. pneumophila* RalF protein fused to GFP (GFP-RalF) as a positive control or LepB. The relative percentage of alkaline phosphatase protein in the tissue culture medium was determined, and is shown as the percentage in tested cells relative to the percentage in control cells producing YFP. Values are the average calculated from three independent assay wells  $\pm$  standard deviation, and show that both GFP-RalF and LepB interfered with alkaline phosphatase secretion significantly, as determined by unpaired two-tailed Student's *t*-test (asterisk,  $P < 0.001$ ).



preferential localization of LepB with the GTP-locked Rab1(Q70L) protein *in vivo*, are results that would be more consistent with LepB functioning as a GAP rather than as an effector of Rab1. Thus, purified LepB was analysed for GAP activity *in vitro* using Rab1 as a substrate. Rab1 GTP hydrolysis was stimulated on the addition of purified LepB (Fig. 4a). The catalytic efficiency ( $k_{\text{cat}}/K_m$ ) of Rab1 GTP hydrolysis stimulated by LepB was approximately  $10^4 \text{ M}^{-1} \text{ s}^{-1}$ , which was similar to the catalytic efficiency measured for yeast Gyp1, a previously characterized eukaryotic GAP that stimulates Rab1 GTP hydrolysis<sup>10,25</sup>. The specificity of the LepB GAP activity was tested using several mammalian Rab proteins as substrates. Rab1 was the only protein in which a significant increase in GTP hydrolysis was observed on the addition of LepB (Fig. 4b). These data show that LepB has GAP activity specific for Rab1, making this the only bacterial protein currently known that has the capacity to modulate Rab function by stimulating GTP hydrolysis.

## Discussion

These data indicate that proteins translocated into host cells by *L. pneumophila* have activities that control three critical events involved in the dynamic cycling of the Rab1 GTPase on membranes. DrrA is a multifunctional enzyme with both GDF and GEF activities.



**Figure 4 | LepB is a Rab1-specific GAP.** **a**, Inorganic phosphate release ( $\Delta A_{360}$ ) resulting from Rab1-GTP hydrolysis was measured over time (min) after the addition of purified His-tagged LepB or yeast Gyp1 at the concentrations indicated, and compared to the intrinsic rate of GTP hydrolysis by Rab1 (Rab1 alone). **b**, Inorganic phosphate release ( $\Delta A_{360}$ ) resulting from GTP hydrolysis was measured over time (min) for the indicated Rab proteins after the addition of 150 nM LepB (red) or 50 nM LepB (green), and compared to the intrinsic rate of GTP hydrolysis (No LepB). Plots show results from two different assay wells for each condition tested.

Previously, the only cellular factor shown to have GDF activity was the eukaryotic Rabac1 protein<sup>21</sup>. Because Rabac1 is an integral membrane protein that seems to have multiple functions in controlling membrane traffic<sup>26</sup>, it has been difficult to demonstrate the specific importance of the GDF activity in recruiting Rab GTPases to subcellular organelles. Our data on the *L. pneumophila* DrrA protein, however, demonstrate clearly that both GDF and GEF activities are needed for membrane recruitment of the Rab1 protein to a plasma-membrane-derived organelle. These findings raise the possibility that eukaryotic proteins that function as Rab GEFs might also have separate domains or subunits that mediate Rab GDI displacement. Thus, these studies not only show the importance of GDF activity in the recruitment of Rab proteins to membranes but also provide new insight into a process that might be used by other Rab GEFs to mediate nucleotide exchange.

In addition to encoding an activator of Rab1, our data demonstrate that the *L. pneumophila* LepB protein is a GAP for Rab1. The ability of LepB to deactivate Rab1 through stimulating GTP hydrolysis could facilitate removal of the GTPase from the early LCV after this plasma-membrane-derived compartment has undergone maturation events mediated by the association and fusion of secretory vesicles. Indeed, we determined that Rab1 association with the LCV membrane was transient and the cycling of Rab1 off the LCV occurred at a time at which LepB began to accumulate on this organelle. Being a large protein with a calculated molecular weight of 148.6 kilodaltons, it is likely that LepB has multiple effector activities, as discovered for the smaller DrrA protein. It will be interesting to locate the GAP domain in LepB and determine whether this region is required for the egress defect reported previously for a *L. pneumophila* lepB mutant<sup>24</sup>. Identifying the GAP domain might be challenging because LepB does not possess the canonical TBC (Tre-2, Bub2 and Cdc16) domain found in most eukaryotic Rab GAPs<sup>9,10</sup>. Although LepB may still use arginine and glutamine residues for a dual-finger mechanism to stabilize the transition state of the GTP hydrolysis reaction of the Rab, the lack of homology to other TBC proteins indicates that the structure of the domain in LepB containing catalytic residues important for GAP activity is likely to be different. Thus, future investigations into the structure and function of LepB are likely to reveal how convergent evolutionary processes have led to the emergence of a bacterial Rab GAP.

On the basis of these findings, we propose the following model for how *L. pneumophila* proteins regulate Rab1 activity on the LCV membrane. DrrA has the intrinsic ability to bind to the cytosolic surface of the plasma membrane<sup>19</sup>. Because the LCV is initially a plasma-membrane-derived organelle, translocated DrrA protein associates with the immature LCV formed on bacterial internalization. The GDF activity of DrrA promotes the sampling of cytosolic Rab proteins associated with Rab GDI. When Rab1-Rab-GDI complexes are sampled, the removal of Rab GDI permits the GEF domain in DrrA to activate Rab1, leading to the stable association of Rab1 with the LCV membrane. Active Rab1 facilitates organization of the LCV membrane to promote the delivery and fusion of endoplasmic-reticulum-derived vesicles<sup>14,27</sup>. DrrA cycles off because the composition of the vacuole membrane changes. The LepB protein accumulates on the LCV membrane and the GAP activity facilitates the removal of Rab1 by stimulating GTP hydrolysis. The removal of Rab1 coincides with fusion of the LCV with the host endoplasmic reticulum and replication of *L. pneumophila* in a vacuole containing endoplasmic reticulum proteins.

## METHODS SUMMARY

Plasmids encoding 3×Flag-Rabs, Rab GDI and DrrA were transfected into HEK293 cells, and 3×Flag-Rab complexes were isolated using anti-Flag M2 affinity gel (Sigma). Recombinant His-tagged DrrA proteins were purified from *Escherichia coli* as described<sup>19</sup>. Rab GDI release was detected by immunoblot analysis using anti-Rab-GDI antibody (Synaptic Systems). For the GEF assays, 3×Flag-Rab1-Rab-GDI complexes were eluted from the anti-Flag agarose

using 3×Flag peptide, and were concentrated using an Amicon centrifugal filter unit with Ultracel-30 membrane (Millipore). The exchange reaction was initiated on addition of the non-hydrolyzable GTP analogue mant-GppNHp, and fluorescence was monitored at 460 nm following excitation at 290 nm using a Sapphire multimode microplate spectrophotometer. Recombinant His–LepB and GST–Rabs were purified from *E. coli* using standard procedures and were used in binding assays as described<sup>28,29</sup>. Rabbits were immunized with purified LepB to generate LepB-specific polyclonal antibodies. LepB co-localization with RFP–Rab1 and GFP–Rab1(Q70L), and LepB-mediated inhibition of alkaline phosphatase secretion, were analysed as described<sup>30</sup>. Bone-marrow macrophages derived from A/J mice were infected with *L. pneumophila*, and LepB, Rab1 and DrrA co-localization with vacuoles containing *L. pneumophila* was determined by fluorescence microscopy as described<sup>19,31</sup>. Purified His–LepB and GTP-loaded GST–Rab proteins were used in the GAP assays with purified yeast Gyp1 as a control. The EnzChek phosphate assay (Molecular Probes) was used to measure GTP hydrolysis as described<sup>25</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 8 September; accepted 4 October 2007.**

**Published online 21 October 2007.**

- Zerial, M. & McBride, H. Rab proteins as membrane organizers. *Nature Rev. Mol. Cell Biol.* **2**, 107–117 (2001).
- Rink, J., Ghigo, E., Kalaidzidis, Y. & Zerial, M. Rab conversion as a mechanism of progression from early to late endosomes. *Cell* **122**, 735–749 (2005).
- Sasaki, T. *et al.* Purification and characterization from bovine brain cytosol of a protein that inhibits the dissociation of GDP from and the subsequent binding of GTP to smg p25A, a ras p21-like GTP-binding protein. *J. Biol. Chem.* **265**, 2333–2337 (1990).
- Araki, S., Kikuchi, A., Hata, Y., Isomura, M. & Takai, Y. Regulation of reversible binding of smg p25A, a ras p21-like GTP-binding protein, to synaptic plasma membranes and vesicles by its specific regulatory protein, GDP dissociation inhibitor. *J. Biol. Chem.* **265**, 13007–13015 (1990).
- Ullrich, O. *et al.* Rab GDP dissociation inhibitor as a general regulator for the membrane association of rab proteins. *J. Biol. Chem.* **268**, 18143–18150 (1993).
- Dirac-Svejstrup, A. B., Sumizawa, T. & Pfeffer, S. R. Identification of a GDI displacement factor that releases endosomal Rab GTPases from Rab-GDI. *EMBO J.* **16**, 465–472 (1997).
- Markgraf, D. F., Peplowska, K. & Ungermann, C. Rab cascades and tethering factors in the endomembrane system. *FEBS Lett.* **581**, 2125–2130 (2007).
- Grosshans, B. L., Ortiz, D. & Novick, P. Rabs and their effectors: achieving specificity in membrane traffic. *Proc. Natl Acad. Sci. USA* **103**, 11821–11827 (2006).
- Strom, M., Vollmer, P., Tan, T. J. & Gallwitz, D. A yeast GTPase-activating protein that interacts specifically with a member of the Ypt/Rab family. *Nature* **361**, 736–739 (1993).
- Albert, S., Will, E. & Gallwitz, D. Identification of the catalytic domains and their functionally critical arginine residues of two yeast GTPase-activating proteins specific for Ypt/Rab transport GTPases. *EMBO J.* **18**, 5216–5225 (1999).
- Horwitz, M. A. Formation of a novel phagosome by the Legionnaires' disease bacterium (*Legionella pneumophila*) in human monocytes. *J. Exp. Med.* **158**, 1319–1331 (1983).
- Swanson, M. S. & Isberg, R. R. Association of *Legionella pneumophila* with the macrophage endoplasmic reticulum. *Infect. Immun.* **63**, 3609–3620 (1995).
- Kagan, J. C. & Roy, C. R. *Legionella* phagosomes intercept vesicular traffic from endoplasmic reticulum exit sites. *Nature Cell Biol.* **4**, 945–954 (2002).
- Kagan, J. C., Stein, M. P., Pypaert, M. & Roy, C. R. *Legionella* subvert the functions of Rab1 and Sec22b to create a replicative organelle. *J. Exp. Med.* **199**, 1201–1211 (2004).
- Derre, I. & Isberg, R. R. *Legionella pneumophila* replication vacuole formation involves rapid recruitment of proteins of the early secretory system. *Infect. Immun.* **72**, 3048–3053 (2004).
- Segal, G., Purcell, M. & Shuman, H. A. Host cell killing and bacterial conjugation require overlapping sets of genes within a 22-kb region of the *Legionella pneumophila* genome. *Proc. Natl Acad. Sci. USA* **95**, 1669–1674 (1998).
- Vogel, J. P., Andrews, H. L., Wong, S. K. & Isberg, R. R. Conjugative transfer by the virulence system of *Legionella pneumophila*. *Science* **279**, 873–876 (1998).
- Nagai, H., Kagan, J. C., Zhu, X., Kahn, R. A. & Roy, C. R. A bacterial guanine nucleotide exchange factor activates ARF on *Legionella* phagosomes. *Science* **295**, 679–682 (2002).
- Murata, T. *et al.* The *Legionella pneumophila* effector protein DrrA is a Rab1 guanine nucleotide-exchange factor. *Nature Cell Biol.* **8**, 971–977 (2006).
- Machner, M. P. & Isberg, R. R. Targeting of host Rab GTPase function by the intravacuolar pathogen *Legionella pneumophila*. *Dev. Cell* **11**, 47–56 (2006).
- Sivars, U., Aivazian, D. & Pfeffer, S. R. Yip3 catalyses the dissociation of endosomal Rab-GDI complexes. *Nature* **425**, 856–859 (2003).
- Stroupe, C. & Brunger, A. T. Crystal structures of a Rab protein in its inactive and active conformations. *J. Mol. Biol.* **304**, 585–598 (2000).
- Pylypenko, O. *et al.* Structure of doubly prenylated Ypt1:GDI complex and the mechanism of GDI-mediated Rab recycling. *EMBO J.* **25**, 13–23 (2006).
- Chen, J. *et al.* *Legionella* effectors that promote nonlytic release from protozoa. *Science* **303**, 1358–1361 (2004).
- Pan, X., Eathiraj, S., Munson, M. & Lambright, D. G. TBC-domain GAPs for Rab GTPases accelerate GTP hydrolysis by a dual-finger mechanism. *Nature* **442**, 303–306 (2006).
- Geng, J., Shin, M. E., Gilbert, P. M., Collins, R. N. & Burd, C. G. *Saccharomyces cerevisiae* Rab-GDI displacement factor ortholog Yip3p forms distinct complexes with the Ypt1 Rab GTPase and the reticulon Rtn1p. *Eukaryot. Cell* **4**, 1166–1174 (2005).
- Robinson, C. G. & Roy, C. R. Attachment and fusion of endoplasmic reticulum with vacuoles containing *Legionella pneumophila*. *Cell. Microbiol.* **8**, 793–805 (2006).
- Satoh, A., Wang, Y., Malsam, J., Beard, M. B. & Warren, G. Golgin-84 is a rab1 binding partner involved in Golgi structure. *Traffic* **4**, 153–161 (2003).
- Christoforidis, S. & Zerial, M. Purification and identification of novel Rab effectors using affinity chromatography. *Methods* **20**, 403–410 (2000).
- Amor, J. C. *et al.* The structure of RalF, an ADP-ribosylation factor guanine nucleotide exchange factor from *Legionella pneumophila*, reveals the presence of a cap over the active site. *J. Biol. Chem.* **280**, 1392–1400 (2005).
- Kagan, J. C., Murata, T. & Roy, C. R. Analysis of Rab1 recruitment to vacuoles containing *Legionella pneumophila*. *Methods Enzymol.* **403**, 71–81 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank G. Warren for his advice on the Rab1 binding studies, and M. P. Stein and L. Chesnel for technical suggestions and advice. This work was supported by the NIH (C.R.R. and D.G.L.) and an NSF Graduate Research Fellowship (A.I.).

**Author Contributions** C.R.R. supervised the project. A.I. conducted all of the experiments in this study with the exception of the GEF assays, which were conducted by A.D. D.G.L. assisted with the GAP assays. All authors contributed to the writing of the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to C.R.R. ([craig.roy@yale.edu](mailto:craig.roy@yale.edu)).



## METHODS

**Isolation of 3×Flag–Rab–Rab–GDI complexes and Rab GDI release assays.** HEK293 cells transfected with plasmids expressing 3×Flag–Rab proteins and Rab GDI were lysed on ice in 64 mM Tris–HCl, pH 8.0, 150 mM NaCl, 8 mM MgCl<sub>2</sub>, 2 mM EDTA, 0.2 mM GDP, 0.1% CHAPS and protease inhibitors. After clearing by spinning at 14,000g for 10 min, lysates were incubated with anti-Flag M2 affinity gel (Sigma) (approximately  $1.5 \times 10^8$  cells per ml of resin), and subsequently the resin was washed with lysis buffer. To assess GDF activity *in vitro*, purified His–DrrA protein was added (0.45 nmol per 20 µl anti-Flag resin in 140 µl total reaction volume) and incubated for 75 min at 25 °C. Soluble fractions were recovered after centrifugation at 5,000g for 30 s. Bound protein was eluted with protein sample buffer at 100 °C (Fig. 1a), 0.1 M glycine, pH 3.5, at 25 °C (Fig. 1b, beads), or 100 µg per ml 3×Flag peptide (Sigma) (Fig. 1d) in volumes equal to the reaction volume. Rab GDI was detected by immunoblot analysis using anti-Rab-GDI antibody (Synaptic Systems).

**Rab1–Rab-GDI GEF assays.** Rab1–Rab-GDI complex (0.8 µM) was incubated in buffer (50 mM Tris, pH 8.0, 150 mM NaCl, 2 mM MgCl<sub>2</sub>, 0.1% CHAPS) with either 0.25 µM full-length DrrA (1–647) or the minimal DrrA GEF domain (451–647). The exchange reaction was initiated on addition of 10 µM mant-GppNHp. Samples were excited at 290 nm, and emission was monitored at 460 nm using a Sapphire multimode microplate spectrophotometer to measure increased sensitized emission accompanying tryptophan to mant fluorescence resonance energy transfer that occurs on mant-GppNHp loading.

**Binding of LepB to GST–Rabs.** GST-tagged and His-tagged proteins were purified from *E. coli* as described<sup>19</sup>. Binding assays were performed based on described methods<sup>28,29</sup>. Human HEK293 cells transiently transfected with plasmid expressing LepB were lysed in 10 mM HEPES, pH 7.4, 150 mM KCl, 5 mM MgCl<sub>2</sub>, 1 mM DTT, 1% Triton X-100 and protease inhibitors. Protein from post-nuclear supernatants (100 µg) was added to purified GST–Rab proteins (approximately 100 µg) immobilized on glutathione sepharose 4B in the presence of 0.2 mM GDP (for GST and GST–Rab1(S25N)) or 0.2 mM GTP (for GST–Rab1, GST–Rab1(Q70L) and GST–Rab7(Q67L)). Alternatively, purified recombinant His–LepB (20 µg) was added to the immobilized GST–Rabs (100 µg). Binding reactions were performed in 250 µl of lysis buffer containing 0.2 mM guanine nucleotide for 14 h at 4 °C. After washing, bound protein was eluted in 10 mM HEPES, pH 7.4, 1.5 M KCl, 20 mM EDTA and 1 mM DTT containing either 5 mM GTP (for GST and GST–Rab1(S25N)) or 5 mM GDP (for GST–Rab1, GST–Rab1(Q70L) and GST–Rab7(Q67L)).

**Eukaryotic expression of LepB.** For localization studies of LepB and GFP–Rab1 in eukaryotic cells, Chinese hamster ovary (CHO) FcγRII cells expressing LepB along with RFP–Rab1 or GFP–Rab1(Q70L) were fixed and stained with antibody against LepB. Cells were subsequently visualized by fluorescence microscopy using a Nikon Eclipse TE2000-S microscope. Digital images were acquired using a Hamamatsu ORCA-ER camera and exported into Adobe Illustrator C2 for the production of figures. Secretion assays were performed as described previously<sup>30</sup>. Data are presented as the measured alkaline phosphatase activity in the cell culture media 7 h after replacement of the media of transfected cells, relative to the cell-associated alkaline phosphatase activity. Displayed in Fig. 3b are the average values from three independent assay wells ± standard deviation.

**Infection and staining of murine bone-marrow-derived macrophages.** Macrophages were derived from bone marrow precursor cells and infected as described<sup>31,32</sup>. After fixation, macrophages were stained for Rab1 and for DrrA as described<sup>19,31</sup>, and for LepB with a rabbit polyclonal antibody raised against purified His–LepB. Samples were visualized by fluorescence microscopy using a Nikon Eclipse TE2000-S microscope. At least 100 LCVs were counted for each protein examined, and the data are representative of two independent assays.

**GAP assays.** GAP assays were performed as described<sup>25</sup>. Purified GST–Rab proteins were loaded with GTP, and unbound nucleotide was removed using a desalting column. Phosphate release was measured in reactions containing 27 µM Rab1 using the EnzChek phosphate assay (Molecular Probes) after the addition of purified LepB–His or yeast Gyp1. Catalytic efficiency of the LepB–His GAP reaction using GST–Rab1 was calculated as described<sup>25</sup>. To compare the activities of LepB–His for the different Rab proteins, initial velocities were calculated: from 810–2,208 s of the reactions with Rab1, Rab5, Rab7 and Rab33; from 973–1,980 s of the reactions with Rab2 and Rab11; and from 3,025–3,966 s of the reaction with Rab6. These rates with and without LepB–His were analysed with the following equation:

$$(k_{\text{cat}}/K_m)/k_{\text{intr}} = ((V_g/V_i) - 1)/[\text{GAP}]$$

Where  $V_g$  is the rate with LepB–His,  $V_i$  is the rate without LepB–His, and [GAP] is the concentration of LepB–His.

**Construction of plasmids and  $\Delta$ lepB mutant strain.** His-tagged versions of LepB were constructed by cloning the gene amplified from *L. pneumophila* genomic DNA into the NdeI and BamHI sites of pET15b, and NdeI and XhoI sites of pET26b. For cloning into pcDNA 4/TO, *lepB* was amplified and inserted into the BamHI and XhoI sites of the digested vector. Full-length complementary DNA encoding human GDP dissociation inhibitor 1 was obtained from Open Biosystems (IHS1380-97430805). For mammalian cell expression, *GDI1* was amplified and inserted into the BamHI and XhoI sites of pcDNA 4/TO. A 3×Flag expression vector was created for tagging Rabs by the Flag epitope into the KpnI and BamHI sites of pcDNA 4/TO. This vector, 3×Flag pcDNA 4/TO, was digested at the BamHI and XhoI sites to insert the cDNA encoding Rab1 and Rab2. To construct the clean deletion of *lepB* in *L. pneumophila*, pSR47S was used<sup>33</sup>. The genomic regions proximal to *lepB* were amplified and inserted into the XbaI and SacI sites of pSR47S. Construction of the isogenic  $\Delta$ lepB mutant in the *L. pneumophila* strain Lp01 by allelic replacement was as described<sup>34</sup>.

32. Celada, A., Gray, P. W., Rinderknecht, E. & Schreiber, R. D. Evidence for a gamma-interferon receptor that regulates macrophage tumoricidal activity. *J. Exp. Med.* **160**, 55–74 (1984).
33. Merriam, J. J., Mathur, R., Maxfield-Boumil, R. & Isberg, R. R. Analysis of the *Legionella pneumophila* *flil* gene: intracellular growth of a defined mutant defective for flagellum biosynthesis. *Infect. Immun.* **65**, 2497–2501 (1997).
34. Campodonico, E. M., Chesnel, L. & Roy, C. R. A yeast genetic system for the identification and characterization of substrate proteins transferred into host cells by the *Legionella pneumophila* Dot/Icm system. *Mol. Microbiol.* **56**, 918–933 (2005).

## ARTICLES

# Portability of paddle motif function and pharmacology in voltage sensors

AbdulRasheed A. Alabi<sup>1\*</sup>, Maria Isabel Bahamonde<sup>1\*</sup>, Hoi Jong Jung<sup>2</sup>, Jae Il Kim<sup>2</sup> & Kenton J. Swartz<sup>1</sup>

**Voltage-sensing domains enable membrane proteins to sense and react to changes in membrane voltage. Although identifiable S1–S4 voltage-sensing domains are found in an array of conventional ion channels and in other membrane proteins that lack pore domains, the extent to which their voltage-sensing mechanisms are conserved is unknown. Here we show that the voltage-sensor paddle, a motif composed of S3b and S4 helices, can drive channel opening with membrane depolarization when transplanted from an archaeobacterial voltage-activated potassium channel (KvAP) or voltage-sensing domain proteins (Hv1 and Ci-VSP) into eukaryotic voltage-activated potassium channels. Tarantula toxins that partition into membranes can interact with these paddle motifs at the protein–lipid interface and similarly perturb voltage-sensor activation in both ion channels and proteins with a voltage-sensing domain. Our results show that paddle motifs are modular, that their functions are conserved in voltage sensors, and that they move in the relatively unconstrained environment of the lipid membrane. The widespread targeting of voltage-sensor paddles by toxins demonstrates that this modular structural motif is an important pharmacological target.**

Ion channels that open and close in response to changes in membrane voltage have a modular architecture, with a central pore domain that determines ion selectivity, and four surrounding voltage-sensing domains that move in response to changes in membrane voltage to drive opening of the pore<sup>1–5</sup> (Fig. 1a). Although X-ray structures have now been solved for two voltage-activated potassium (Kv) channels<sup>1,6–9</sup>, the structural basis of voltage sensing remains controversial<sup>10–12</sup>. A seminal observation in the X-ray structures of the KvAP channel, an archaeobacterial Kv channel from *Aeropyrum pernix*, was that the S3b helix and the charge-bearing S4 helix within the voltage-sensing domain form a helix–turn–helix structure, termed the paddle motif<sup>1,8,9</sup>. Studies on KvAP<sup>1,9,13–16</sup> indicate that this voltage-sensor paddle is buried in the membrane and that it moves at the protein–lipid interface; this contrasts with models for eukaryotic Kv channels, in which the S4 helix is protected from membrane lipids by other regions of the protein<sup>10–12,17–21</sup>. Voltage-sensing domains have also been described recently in voltage-sensing proteins that lack associated pore domains<sup>5,22,23</sup>. In the *Ciona intestinalis* voltage sensitive phosphatase (Ci-VSP) the voltage-sensing domain is coupled to a phosphatase domain, and in the voltage-activated proton channel (Hv1) the voltage-sensing domain itself is thought to function as a proton channel. Here we explore whether the mechanisms of voltage sensing are conserved between the distantly related eukaryotic and archaeobacterial Kv channels, and the newly discovered voltage-sensing-domain proteins Ci-VSP and Hv1. Using a chimaera approach, we first examine whether specific structural elements within voltage sensors can be transferred between Kv channels and voltage-sensing-domain proteins while preserving functional responses to changes in membrane voltage. We then use a family of tarantula toxins that interact with voltage-sensing domains to explore the structural integrity of these modular motifs and their disposition with respect to the lipid membrane.

## Chimaeras between Kv channels

We began by generating chimaeras between the archaeobacterial KvAP channel<sup>24</sup> and the eukaryotic Kv2.1 channel from rat brain<sup>25</sup> to define

the interchangeable regions. Transfer of the KvAP pore domain into Kv2.1 results in channels that open in response to membrane depolarization (Fig. 1a, Supplementary Fig. 1 and Supplementary Table 1), as long as the S4–S5 linker helix and the most carboxy-terminal region of the S6 helix are from the same channel—a result that reaffirms the modular architecture of Kv channels and the important roles of the S4–S5 and S6 helices in coupling the voltage-sensing and pore domains<sup>4,26</sup>. We then turned to identifying the regions within the voltage-sensing domains that are compatible, and in doing so produced >60 chimaeras that varied in the amount of KvAP sequence they contained and in the region transferred (Fig. 1 and Supplementary Fig. 1). Most of these constructs result in non-functional channels (Fig. 1a and Supplementary Fig. 1), a result that is not surprising given that these constructs typically contain many (often radical) amino acid changes. One region stands out, however, where relatively large portions of KvAP can be transferred into Kv2.1 without disrupting channel function. The transferable region begins at the junction between S3a and S3b helices and ends just past the first four critical Arg residues in S4 (refs 17, 27 and 28), corresponding to the paddle motif identified in the X-ray structure of KvAP<sup>1</sup> (Fig. 1a–d). These paddle chimaeras display robust voltage-activated K<sup>+</sup> currents and gating properties that are qualitatively similar to Kv2.1 (Fig. 1d). A similar chimaera in which the paddle of KvAP is transferred to the Shaker Kv channel from *Drosophila*<sup>29</sup> also results in functional Kv channels (Fig. 2a and Supplementary Table 1), indicating that the results with Kv2.1 are applicable to other types of Kv channels. Extending the region transferred by extension on the amino-terminal side of the paddle into the S3a helix or on the carboxy-terminal side beyond the first four Arg residues in S4 results in non-functional channels (Fig. 1b and Supplementary Fig. 1). The inability to extend into the C-terminal portion of S4 is consistent with the presence of crucial protein–protein interactions between the inner regions of the S4 and S5 helices<sup>30</sup>. The preservation of channel function observed in the paddle chimaeras is quite remarkable considering the large number of amino acid substitutions they contain. One such chimaera (C7[S3–S4] AP, using the nomenclature Cx[region transferred]donor channel,

<sup>1</sup>Molecular Physiology and Biophysics Section, Porter Neuroscience Research Center, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>2</sup>Department of Life Sciences, Gwangju Institute of Science and Technology, Gwangju, 500-712, Korea.

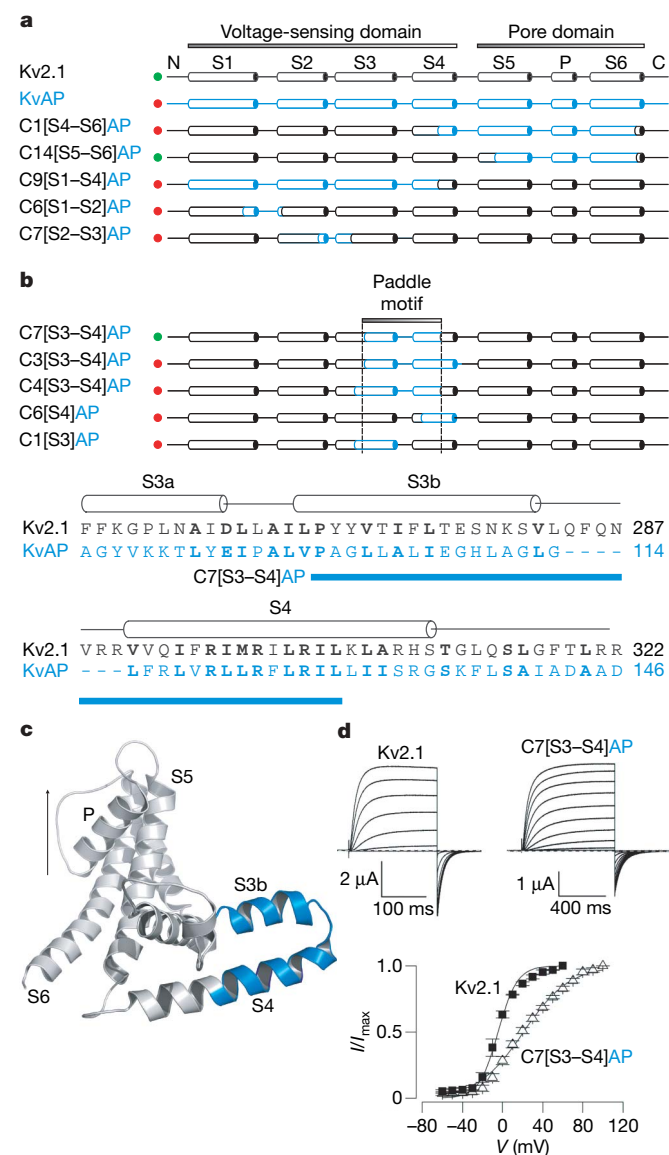
\*These authors contributed equally to this work.

where x is the chimaera number within that region) contains 25 single-residue substitutions (15 of which are non-conservative) and a 7-residue deletion. In the context of so many chimaeras that do not result in functional channels, the successful transfer of the paddle motif indicates that this region is modular and unique in its paucity of rigidly constraining side-chain interactions with other parts of the protein.

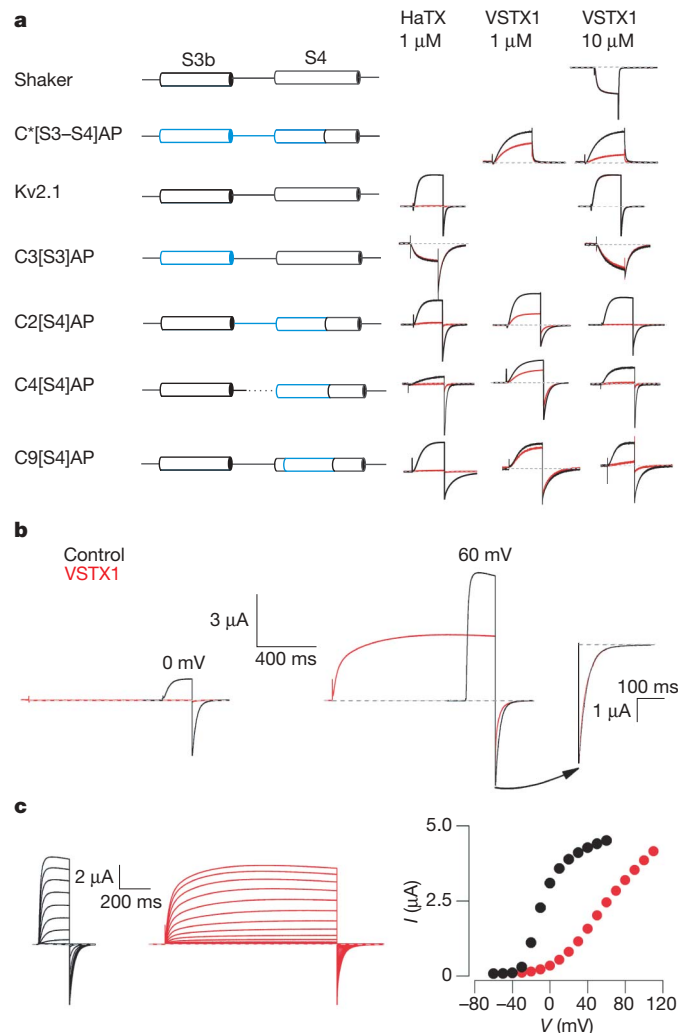
### Toxins interacting with paddles in Kv channels

To further explore the structural and functional integrity of the chimaeras, we examined their sensitivities to tarantula toxins known to

inhibit Kv channels by interacting with voltage sensors. The two toxins we focused on are hanatoxin (HaTX), which does not interact with KvAP but inhibits Kv2.1 by interacting with its voltage-sensor paddle<sup>31–37</sup>, and voltage-sensor toxin 1 (VSTX1), a related tarantula toxin that does not inhibit either Kv2.1 or Shaker (Fig. 2a) but inhibits KvAP by binding somewhere within its S1–S4 domain<sup>9,24,38</sup>. Transferring the KvAP paddle into Shaker (C\*[S3–S4]AP) renders



**Figure 1 | Transfer of the voltage-sensor paddle motif from KvAP to Kv2.1 channels.** **a**, Overview of chimaeras between KvAP (blue) and Kv2.1 (black). Constructs that result in functional Kv-channel activity when expressed in oocytes are indicated with green circles and those that are non-functional are indicated with red circles. **b**, Definition of the region within the voltage sensor of KvAP (indicated by dashed lines) that results in functional channels when transferred to Kv2.1. Alignment between KvAP and Kv2.1 in the region from S3 to S4, highlighting the stretch of residues transferred to form C7[S3–S4]AP (blue bar). Conserved residues are shown in bold lettering. **c**, Backbone fold of a single subunit of KvAP (left, PDB accession number, 2A0L), depicting the paddle region in blue. The arrow indicates the permeation pathway for potassium ions. These and all subsequent structures were created using PyMOL (DeLano Scientific LLC). **d**, Potassium currents and tail-current voltage-activation relationship ( $n = 5–12$ , error bars represent s.e.m.) for Kv2.1 and the C7[S3–S4]AP paddle chimaera after expression in oocytes. The holding voltage was  $-80$  mV and the tail voltage was  $-60$  mV.



**Figure 2 | Sensitivity of KvAP paddle chimaeras to extracellular tarantula toxins.** **a**, Channel constructs, designated at the left with KvAP segments shown in blue, were expressed in oocytes and potassium currents elicited by depolarizations in the absence (black) or presence (red) of either HaTX or VSTX1. Depolarizations were to voltages near the foot of the voltage-activation relationship (relative open probability of  $<0.3$ ) for each construct. All chimaeras involving Kv2.1 are defined in Supplementary Fig. 1. The paddle chimaera in Shaker (C\*[S3–S4]AP) was generated by transplanting Pro 99–Arg 126 of KvAP into Pro 322–Arg 371 of Shaker. C\*[S3–S4]AP was studied with a low  $K^+$  external solution, and all others were studied with a high  $K^+$  external solution (see Methods). Shaker has a very low sensitivity to HaTX<sup>48</sup> and was not studied. VSTX1-insensitive channels were only studied at the highest VSTX1 concentration. **b**, VSTX1 inhibition of chimaera C2[S4]AP is voltage-dependent. Potassium currents were recorded for weak (0 mV, left) and strong (60 mV, right) depolarizations, before and after addition of 12  $\mu$ M VSTX1. The inset to the right shows scaled tail currents after depolarization to 60 mV. **c**, Families of currents recorded in response to depolarizations in the absence (black) and presence (red) of 12  $\mu$ M VSTX1. The holding voltage was  $-90$  mV and the tail voltage was  $-60$  mV. Corresponding tail-current voltage-activation relationships for the traces shown, in which tail-current amplitude is plotted against test voltage.



this eukaryotic Kv channel sensitive to extracellular VSTX1 (Fig. 2a), indicating that this tarantula toxin interacts with the paddle motif in KvAP, similar to the interaction of HaTX with the Kv2.1 channel. Transferring the S3b helix of KvAP alone into Kv2.1 (C3[S3]AP) results in a channel that is insensitive to extracellular HaTX (Fig. 2a); this makes sense because KvAP is insensitive to HaTX, and the most crucial determinants of HaTX binding to Kv2.1 are localized within S3b<sup>31–37</sup>. In the case of VSTX1, transferring the S4 helix alone from KvAP into Kv2.1 (C2[S4]AP, C4[S4]AP, C9[S4]AP) is sufficient to render the recipient channel sensitive to VSTX1, in each case without disrupting sensitivity to HaTX (Fig. 2a). Although these results do not preclude important interactions between VSTX1 and the S3b helix of the paddle, they point to a particularly important role of the S4 helix.

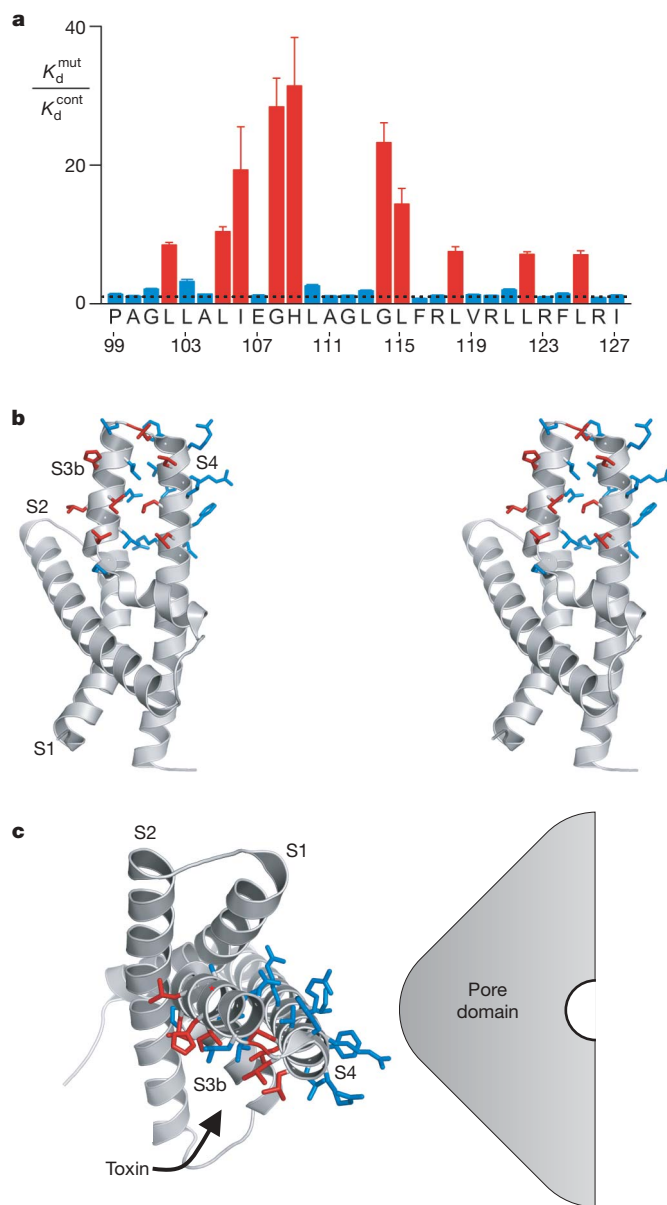
One of the hallmarks of toxins that inhibit by stabilizing the resting state of the paddle motif (for example, HaTX) is that inhibition can be overcome when the channel is activated by strong depolarization of the membrane<sup>31,35,36</sup>. Although VSTX1 prevents opening of chimaeras containing the S4 helix from KvAP in the Kv2.1 channel (for example, C2[S4]AP) when activating the channel using weak depolarizations to 0 mV, the channel opens robustly in response to large depolarizations to +60 mV (Fig. 2b). In effect, VSTX1 shifts activation of the channel to more depolarized voltages (Fig. 2c), similar to what is seen in the case of HaTX<sup>31,35,36</sup>. Similar shifts of channel activation are observed for each of the VSTX1-sensitive paddle chimaeras studied here (data not shown). The kinetics of channel activation are biphasic in the presence of VSTX1 (Fig. 2b), raising the possibility that the toxin unbinds during strong depolarizations—an inference that is supported by experiments using multiple-pulse protocols to measure toxin dissociation (Supplementary Fig. 2). Taken together, these results indicate that tarantula toxins interact with helices in the voltage-sensor paddle motif and that these interactions are appropriately preserved in the chimaeras. Both of these toxins are thought to interact with their target channels within the lipid membrane<sup>14,36,37,39,40</sup>, which fits nicely with the portability of the paddle between the two types of distantly related Kv channels because it suggests that the unconstrained environment is actually the lipid membrane.

### Structural analysis of the toxin receptor

Although it is clear that the paddle motif is the receptor for tarantula toxins such as HaTX in Kv2.1<sup>31–37</sup>, the disposition of the toxin-binding surface with respect to the lipid membrane and other regions of the protein is unclear. HaTX does not interact tightly with either KvAP or Kv1.2—the two Kv channels for which X-ray structures are available—and the sequence similarity between these channels and Kv2.1 is quite low, precluding assignment of crucial residues within the existing structures. The interaction of VSTX1 with chimaeras containing the entire paddle of KvAP (Fig. 2), however, offers an opportunity to explore the structure of a tarantula toxin receptor because there are three X-ray structures of KvAP<sup>1,8</sup>, and the structure of the paddle motif in these is very similar (whether or not an antibody is bound).

To define the critical residues within the paddle motif that may interact with VSTX1, we Ala-scanned the paddle region of the C\*[S3–S4]AP chimaera between KvAP and Shaker. Most of the 29 residues within the KvAP paddle in this construct were individually mutated to Ala (except for native Ala residues, which were mutated to Val), and the effects on the apparent affinity of the toxin were determined<sup>31–37</sup>. Ten mutants stand out as having dramatic effects on toxin sensitivity: L102A, L105A, I106A, G108A and H109A within the S3b helix, G114A and L115A within the connecting loop, and L118A, L122A and L125A within the S4 helix (Fig. 3a and Supplementary Table 2). When mapped onto the high-resolution structure of the isolated voltage-sensing domain of KvAP<sup>1</sup>, the results provide a remarkable picture of how the toxin must dock onto the paddle motif because, with the exception of G108A, all of the crucial side chains are

located on a contiguous surface of the S3b–S4 helix–turn–helix motif (Fig. 3b). This surface is striking in that it is comprised largely of the aliphatic residues Ile and Leu, with His 109 being the single important polar residue. We can see the disposition of this surface with respect to the rest of the channel and the surrounding lipid membrane by positioning the voltage-sensing domain of KvAP relative to the pore domain in a manner that is consistent with the X-ray structure of Kv1.2 (ref. 7, Fig. 3c). The surface of the paddle that is critical for interacting with VSTX1 projects out towards the surrounding



**Figure 3 | Structural analysis of the toxin–paddle interaction.** **a**, Shown is an Ala-scan of the paddle motif of KvAP in the C\*[S3–S4]AP chimaera, where perturbations in apparent VSTX1 affinity ( $K_d^{\text{mut}}/K_d^{\text{cont}}$ ) are plotted for individual Ala and Val mutants. The dashed line marks a value of 1 and numbering corresponds to the amino acid sequence of KvAP. Each mutant was examined initially using a concentration of toxin near the  $K_d$  for the control chimaera; mutants displaying higher  $K_d$  values were further examined using higher toxin concentrations.  $n = 3–5$  for each toxin concentration, and error bars are s.e.m. **b**, Stereo pair of the isolated voltage-sensing domain of KvAP (PDB accession number, 1ORS), with side chains in the paddle coloured according to perturbations in toxin affinity as in **a**. **c**, Positioning of the voltage-sensing domain of KvAP adjacent to a hypothetical pore domain according to the X-ray structure of Kv1.2 (PDB accession number, 2A79). The  $\alpha$ -carbon of Gly 108 is indicated by a red asterisk.

lipid bilayer, revealing that the toxin docks onto the paddle motif at the perimeter of the channel. This structural picture is consistent with functional studies, indicating that tarantula toxins interact with Kv channels within the lipid bilayer<sup>14,36,37,39,40</sup>, and that multiple toxins (perhaps up to four) can bind to a single Kv channel relatively independently<sup>31,35</sup>. The intimate involvement of membrane lipids in contributing to the toxin receptor<sup>40</sup> is also supported by these mutagenesis results, because non-annular lipids that have been resolved in crystal structures<sup>41</sup> often intercalate between the types of aliphatic residues dominating the surface of the paddle that is crucial for toxin binding.

### Chimaeras with paddles from voltage-sensing proteins

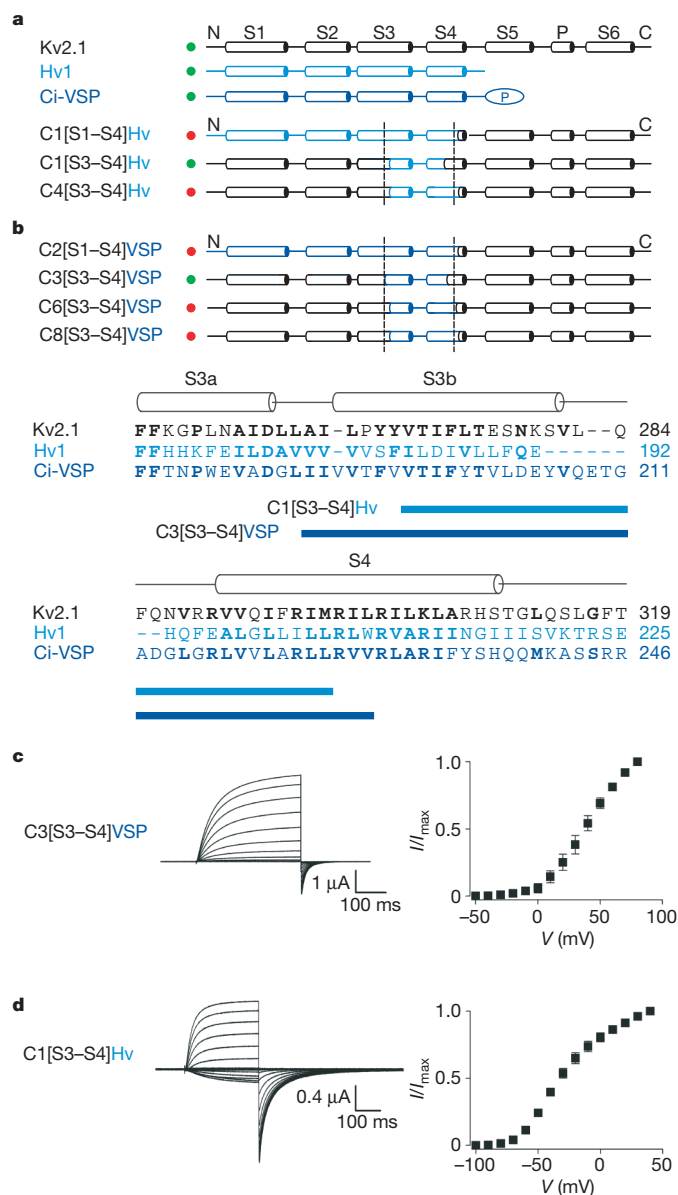
Although the recently discovered voltage-sensing proteins, Ci-VSP and Hv1<sup>5,22,23</sup>, contain recognizable S1–S4 transmembrane segments, the relationship between their voltage-sensing domains and those found in ion channels has yet to be explored. Emboldened by the successful transfer of paddle motifs between KvAP and eukaryotic Kv channels, we examined whether similar regions of the voltage-sensing proteins can drive opening of eukaryotic Kv channels. We constructed a series of chimaeras by transferring regions of the voltage-sensing domains of Ci-VSP and Hv1 into Kv2.1, and, similar to our results with KvAP, discovered that the paddle region of these voltage-sensing proteins is competent to drive opening of the eukaryotic Kv channel (Fig. 4a–c). Robust voltage-activated K<sup>+</sup> currents are observed for several chimaeras containing the paddle region of the voltage-sensing proteins, but extension of these constructs to include larger regions of the voltage-sensing domains does not result in functional channels (Supplementary Fig. 4 and Supplementary Table 3). These results reinforce the notion that the paddle motif moves in a relatively unconstrained environment and indicate that the general mechanisms of voltage-sensing are similar between Kv channels and proteins with a voltage-sensing domain.

### Toxins interacting with voltage-sensing proteins

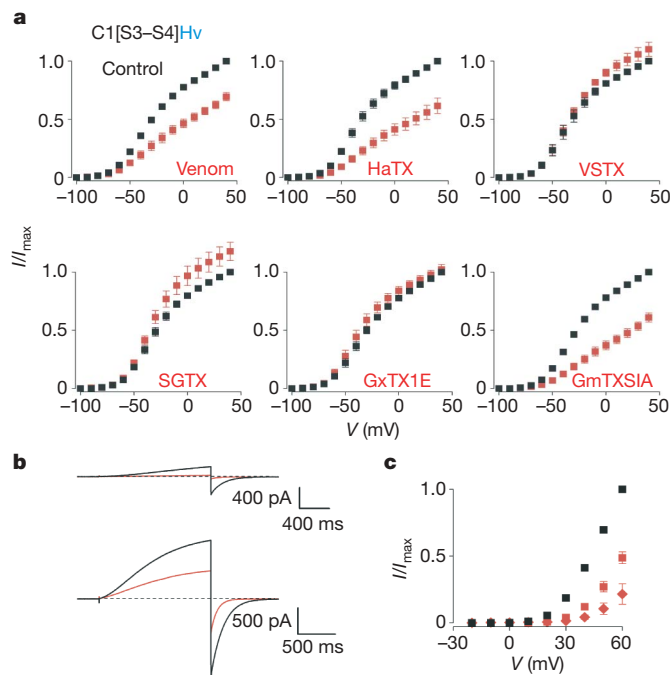
To explore whether tarantula toxins can interact with the paddle motifs of proteins with a voltage-sensing domain, we tested the activity of five tarantula toxins and crude *Grammostola spatulata* venom on chimaeras containing paddles from Hv1 or Ci-VSP in the Kv2.1 channel. Remarkably, HaTX, grammatxin-SIA (GmTXSIA)<sup>3,42</sup> and crude venom inhibit the Hv1 paddle chimaera by shifting activation to positive voltages (Fig. 5a), similar to what is seen for the interaction of tarantula toxins with Kv channels<sup>37</sup>. To confirm that HaTX interacts with the paddle region of Hv1, we mutated several residues within the transferred region, and found that the D185A mutant effectively eliminates the inhibitory effects of the toxin (Supplementary Fig. 5). To examine whether the interaction between HaTX and the Hv1 paddle in the chimaera is predictive of interactions between the toxin and the wild-type Hv1 protein, we expressed Hv1 alone and tested whether HaTX can inhibit voltage-activated proton currents. Extracellular application of HaTX produces robust inhibition of Hv1 proton currents (Fig. 5b), revealing that the toxin can interact with the voltage-sensing domain of the Hv1 protein itself. When examined over a range of membrane voltages, we observed that HaTX shifts activation of the Hv1 protein to more positive voltages (Fig. 5b, c). Taken together, these results indicate that tarantula toxins can interact with paddle motifs, regardless of whether they are present in Kv channels or proteins with a voltage-sensing domain. Although HaTX, *Scodra griseipes* toxin 1 (SGTX1) (ref. 43), VSTX1 (ref. 24), GmTXSIA and guangxitoxin 1E (GxTX1E)<sup>44</sup> have little effect on the activity of the Ci-VSP chimaera, the crude venom contains inhibitory activity (see Supplementary Fig. 6). All of the toxins isolated so far from *G. spatulata* venom are related to HaTX and VSTX1<sup>37</sup>, raising the possibility that tarantula venom contains a related toxin that targets the paddle motif of Ci-VSP.

### Discussion

The present results with paddle chimaeras and tarantula toxins have three general implications for the structural basis of voltage sensing. First, they indicate that the paddle motif is a modular unit that is common to voltage sensors, whether they are found in eukaryotic or archaeobacterial Kv channels, or in voltage-sensing proteins such as the Ci-VSP voltage-sensitive phosphatase or the Hv1 voltage-gated



**Figure 4 | Transfer of the voltage-sensor paddle motif from Hv1 or Ci-VSP into Kv2.1 channels.** **a**, **b**, Overview of chimaeras between Hv1 (**a**, light blue), Ci-VSP (**b**, dark blue) and Kv2.1 (black). Constructs that result in functional Kv channel activity when expressed in oocytes are indicated with green circles and those that are non-functional are indicated with red circles. Dashed lines represent the same as in Fig. 1. The amino acid alignment shows the sequence of Kv2.1, Hv1 and Ci-VSP in S3 to S4, highlighting (blue bars) the stretch of residues transferred to form the two chimaeras indicated. **c**, Current traces and tail-current voltage-activation relationships ( $n = 3$ , error bars are s.e.m.) for a chimaera expressed in oocytes where the paddle of Ci-VSP was transferred into Kv2.1. Test depolarizations were to voltages between  $-50$  mV and  $+80$  mV, holding voltage was  $-80$  mV and tail voltage was  $-50$  mV. **d**, Current traces and tail-current voltage-activation relationships ( $n = 3$ , error bars are s.e.m.) for a chimaera where the paddle of Hv1 was transferred into Kv2.1. Test depolarizations were to voltages between  $-100$  mV and  $+40$  mV, holding voltage was  $-90$  mV and tail voltage was  $-90$  mV.

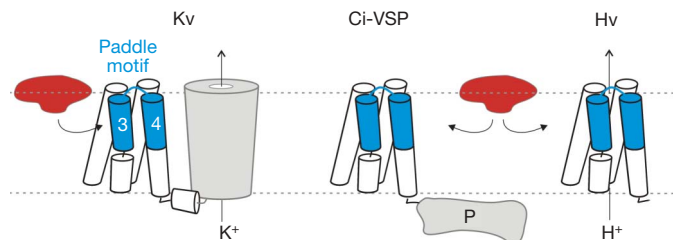


**Figure 5 | Sensitivity of a Hv1 paddle chimaera and the Hv1 proton channel to tarantula toxins.** **a**, Voltage-activation relationships in the absence (black) and presence (red) of tarantula venom or toxins for a chimaera containing the paddle of Hv1 in the Kv2.1 channel. Potassium currents were recorded using 300–500 ms test depolarizations from holding voltages between  $-100$  mV and  $-90$  mV, and tail voltages between  $-100$  mV and  $-90$  mV. Venom was applied at a 1:5,000 dilution and toxin concentrations were  $2 \mu\text{M}$  for HaTX,  $8 \mu\text{M}$  for VSTX1,  $4 \mu\text{M}$  for SGTX1,  $1 \mu\text{M}$  for GxTX1E and  $10 \mu\text{M}$  for GmTXSIA. **b**, Proton currents recorded for Hv1 in response to weak ( $+30$  mV, top) and strong ( $+60$  mV, bottom) depolarizations, both in the absence (black) and presence (red) of  $4 \mu\text{M}$  HaTX. Hv1 was expressed in HEK cells. The holding voltage was  $-40$  mV and tail voltage was  $-60$  mV. **c**, Tail-current voltage-activation relationships for Hv1 recorded in the absence (black) and presence of  $1 \mu\text{M}$  (red squares) or  $4 \mu\text{M}$  (red diamonds) HaTX. For all voltage-activation relationships,  $n = 4$ – $5$  and error bars are s.e.m.

proton channel. The ability of the paddle to support both voltage and toxin sensitivity argues that the fundamental voltage-sensing mechanism is probably conserved between these distantly related membrane proteins.

Second, the portability of the paddle motif amongst voltage sensors that have a low sequence similarity indicates that this motif resides in a relatively unconstrained environment. To successfully transplant such dissimilar paddle motifs in the context of extensive packing with the surrounding protein, one would have to transplant all interacting regions. The ability of tarantula toxins to interact with the paddle within the membrane<sup>14,36,37,39,40</sup>, as well as the projection of residues important for toxin binding out towards the surrounding lipid membrane (Fig. 3), suggests strongly that the lipid membrane is the flexible environment surrounding the paddle motif. Our results do not preclude crucial side-chain interactions between a small subset of residues in the paddle and other parts of the voltage sensor. The charge-carrying Arg residues, for example, are thought to interact with acidic residues in S2 and S3 (refs 28 and 45), and there seems to be specificity in the positions in S4 where charged moieties contribute to the total gating charge<sup>17</sup>. The Arg residues are the most conserved positions in our paddle constructs, so important interactions involving these residues are fully compatible with our results.

Third, the picture of tarantula toxins interacting with crucial residues in both S3b and S4 (Figs 2 and 3), when taken together with the fact that the paddle can move with the toxin continuously bound<sup>36</sup> (Supplementary Fig. 2), indicates that these two helices within the



**Figure 6 | Tarantula toxins interacting with voltage-sensor paddle motifs.** Cartoon depicting the interaction of tarantula toxins with voltage-sensor paddle motifs within the lipid membrane for Kv channels, and for the voltage-sensing-domain proteins, Ci-VSP and Hv1. Only one of the four voltage-sensing domains is shown surrounding the pore domain in Kv channels. P is the phosphatase domain of Ci-VSP.

paddle move roughly as a unit in response to changes in membrane voltage. Large rotations of the S4 helix relative to S3b<sup>18</sup>, for example, would be hard to reconcile with the present picture of tarantula toxins docking onto the paddle motif (Fig. 3).

The demonstration that paddle motifs are portable modules also has exciting implications for studying proteins that contain voltage sensors, because it means that new proteins can be engineered with specific paddle motifs that are tailored for a particular purpose. We used the modular feature of voltage-sensor paddles to screen paddles from proteins with a voltage-sensing domain for interactions with tarantula toxins using Kv channel activity as an assay, which could be adapted for high-throughput screening of chemical libraries for compounds that interact with paddle motifs from any protein that contains a voltage sensor. X-ray structures are available for only KvAP and Kv1.2 channels<sup>1,7,8</sup>, and in each of these structures there are key regions that are poorly defined or distorted. In ref. 46, the portability of the paddle motif is used to solve a new X-ray structure of a Kv channel with a paddle variant, leading to important new insights into the structural basis of voltage sensing.

Previous studies on the interaction of HaTX with voltage sensors in Kv2.1 channels indicate that the paddle motif observed in the KvAP X-ray structure is the key region of the channel that the toxin targets<sup>3,31–37,43</sup>. The present studies with tarantula toxins that interact with paddle motifs in a wide range of distantly related proteins with voltage sensors, including KvAP, Ci-VSP and Hv1 (Fig. 6), establish the general principle that the modular paddle motif is an important pharmacological target. VSTX1 interacts with the paddle motif in KvAP, HaTX can recognize the paddle motif in either Hv1 or Kv2.1, and other toxins in tarantula venom probably interact with the paddle of the voltage-sensitive phosphatase. It will be fascinating to use these emerging ideas about paddle motifs to identify new molecules and drugs that modulate the activity of the large and diverse variety of membrane proteins that contain voltage-sensing domains. It will also be exciting to explore the possibility that ion channels only distantly related to Kv channels—for example, transient receptor potential (TRP) channels—will contain paddle-like motifs that have related roles in channel gating and represent important pharmacological targets. Indeed, several recently discovered tarantula toxins that are very similar to those studied here have been shown to activate the capsaicin receptor channel TRPV1 (ref. 47). It seems that nature has targeted the paddle throughout evolution, with tarantulas being just one example of the organisms that produce paddle toxins<sup>37</sup>. Such widespread targeting of the paddle motif can be explained by the observation that this modular structural motif is a uniquely mobile region within S1–S4 voltage-sensing domains<sup>13</sup>.

## METHODS SUMMARY

Chimaeras and point mutations were generated using KvAP<sup>24</sup>, Ci-VSP<sup>5</sup>, Hv1 (ref. 23), Kv2.1A7 (refs 3 and 25) or Shaker H4 (ref. 29). Most channel constructs were expressed in *Xenopus* oocytes<sup>35</sup> and studied using two-electrode voltage-clamp recording techniques. For most experiments, the external recording



solution contained: 50 mM KCl, 50 mM NaCl, 10–20 mM HEPES, 1 mM MgCl<sub>2</sub> and 0.3 mM CaCl<sub>2</sub>, pH 7.4–7.6 with NaOH. To record outward tail currents for the C\*[S3–S4]AP chimera, the external solution contained: 4 mM KCl, 96 mM NaCl, 10 mM HEPES, 1 mM MgCl<sub>2</sub> and 0.3 mM CaCl<sub>2</sub>, pH 7.6 with NaOH. Macroscopic proton currents were recorded in whole-cell mode using a patch-clamp amplifier after expression of Hv1 in HEK cells. The intracellular (pipette) solution contained: 100 mM Bis-Tris, 75 mM NaCl, 1 mM EGTA and 2 mM MgCl<sub>2</sub>, pH 6.5; the extracellular (bath) solution contained: 100 mM Bis-Tris, 75 mM NaCl, 1 mM EGTA and 3 mM CaCl<sub>2</sub>, pH 6.5.

Hanatoxin was purified from *G. spatulata* venom (Spider Pharm) as described previously<sup>48</sup>. VSTX1, SGTX1 and GxTX1E were synthesized using solid-phase chemical methods, and were folded and purified as described previously<sup>39,43</sup>. GmTXSIA was provided by R. A. Keith and R. A. Lampe.

Voltage-activation relationships for all constructs were obtained by measuring tail currents after a series of membrane depolarizations or by measuring steady-state currents and calculating conductance. Occupancy of closed or resting channels by tarantula toxins was examined using negative holding voltages where open probability is very low, and the fraction of unbound channels ( $F_u$ ) was estimated using depolarizations that are too weak to open toxin-bound channels, as described previously<sup>3,31–37,43</sup> (Supplementary Fig. 3). Example traces showing the inhibitory activity of tarantula toxins were taken for relatively weak depolarizations for that particular channel construct. The apparent equilibrium dissociation constant ( $K_d$ ) was calculated assuming four independent toxin-binding sites per channel, with single occupancy being sufficient to inhibit opening in response to weak depolarizations:

$$K_d = ((1/(1 - F_u^{1/4})) - 1)[\text{Toxin}]$$

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 31 July; accepted 17 September 2007.**

- Jiang, Y. *et al.* X-ray structure of a voltage-dependent K<sup>+</sup> channel. *Nature* **423**, 33–41 (2003).
- Kubo, Y., Baldwin, T. J., Jan, Y. N. & Jan, L. Y. Primary structure and functional expression of a mouse inward rectifier potassium channel. *Nature* **362**, 127–133 (1993).
- Li-Smerin, Y. & Swartz, K. J. Gating modifier toxins reveal a conserved structural motif in voltage-gated Ca<sup>2+</sup> and K<sup>+</sup> channels. *Proc. Natl Acad. Sci. USA* **95**, 8585–8589 (1998).
- Lu, Z., Klem, A. M. & Ramu, Y. Ion conduction pore is conserved among potassium channels. *Nature* **413**, 809–813 (2001).
- Murata, Y., Iwasaki, H., Sasaki, M., Inaba, K. & Okamura, Y. Phosphoinositide phosphatase activity coupled to an intrinsic voltage sensor. *Nature* **435**, 1239–1243 (2005).
- Long, S. B., Campbell, E. B. & MacKinnon, R. Voltage sensor of Kv1.2: structural basis of electromechanical coupling. *Science* **309**, 903–908 (2005).
- Long, S. B., Campbell, E. B. & MacKinnon, R. Crystal structure of a mammalian voltage-dependent Shaker family K<sup>+</sup> channel. *Science* **309**, 897–903 (2005).
- Lee, S. Y., Lee, A., Chen, J. & MacKinnon, R. Structure of the KvAP voltage-dependent K<sup>+</sup> channel and its dependence on the lipid membrane. *Proc. Natl Acad. Sci. USA* **102**, 15441–15446 (2005).
- Jiang, Y., Ruta, V., Chen, J., Lee, A. & MacKinnon, R. The principle of gating charge movement in a voltage-dependent K<sup>+</sup> channel. *Nature* **423**, 42–48 (2003).
- Tombola, F., Pathak, M. M. & Isacoff, E. Y. How far will you go to sense voltage? *Neuron* **48**, 719–725 (2005).
- Ahern, C. A. & Horn, R. Stirring up controversy with a voltage sensor paddle. *Trends Neurosci.* **27**, 303–307 (2004).
- Swartz, K. J. Towards a structural view of gating in potassium channels. *Nature Rev. Neurosci.* **5**, 905–916 (2004).
- Ruta, V., Chen, J. & MacKinnon, R. Calibrated measurement of gating-charge arginine displacement in the KvAP voltage-dependent K<sup>+</sup> channel. *Cell* **123**, 463–475 (2005).
- Lee, S. Y. & MacKinnon, R. A membrane-access mechanism of ion channel inhibition by voltage sensor toxins from spider venom. *Nature* **430**, 232–235 (2004).
- Schmidt, D., Jiang, Q. X. & MacKinnon, R. Phospholipids and the origin of cationic gating charges in voltage sensors. *Nature* **444**, 775–779 (2006).
- Cuello, L. G., Cortes, D. M. & Perozo, E. Molecular architecture of the KvAP voltage-dependent K<sup>+</sup> channel in a lipid bilayer. *Science* **306**, 491–495 (2004).
- Ahern, C. A. & Horn, R. Specificity of charge-carrying residues in the voltage sensor of potassium channels. *J. Gen. Physiol.* **123**, 205–216 (2004).
- Campos, F. V., Chanda, B., Roux, B. & Bezanilla, F. Two atomic constraints unambiguously position the S4 segment relative to S1 and S2 segments in the closed state of Shaker K channel. *Proc. Natl Acad. Sci. USA* **104**, 7904–7909 (2007).
- Grabe, M., Lai, H. C., Jain, M., Nung Jan, Y. & Yeh Jan, L. Structure prediction for the down state of a potassium channel voltage sensor. *Nature* **445**, 550–553 (2007).
- Tombola, F., Pathak, M. M., Gorostiza, P. & Isacoff, E. Y. The twisted ion-permeation pathway of a resting voltage-sensing domain. *Nature* **445**, 546–549 (2007).
- Chanda, B., Asamoah, O. K., Blunck, R., Roux, B. & Bezanilla, F. Gating charge displacement in voltage-gated ion channels involves limited transmembrane movement. *Nature* **436**, 852–856 (2005).
- Sasaki, M., Takagi, M. & Okamura, Y. A voltage sensor-domain protein is a voltage-gated proton channel. *Science* **312**, 589–592 (2006).
- Ramsey, I. S., Moran, M. M., Chong, J. A. & Clapham, D. E. A voltage-gated proton-selective channel lacking the pore domain. *Nature* **440**, 1213–1216 (2006).
- Ruta, V., Jiang, Y., Lee, A., Chen, J. & MacKinnon, R. Functional analysis of an archaebacterial voltage-dependent K<sup>+</sup> channel. *Nature* **422**, 180–185 (2003).
- Frech, G. C., VanDongen, A. M., Schuster, G., Brown, A. M. & Joho, R. H. A novel potassium channel with delayed rectifier properties isolated from rat brain by expression cloning. *Nature* **340**, 642–645 (1989).
- Lu, Z., Klem, A. M. & Ramu, Y. Coupling between voltage sensors and activation gate in voltage-gated K<sup>+</sup> channels. *J. Gen. Physiol.* **120**, 663–676 (2002).
- Aggarwal, S. K. & MacKinnon, R. Contribution of the S4 segment to gating charge in the Shaker K<sup>+</sup> channel. *Neuron* **16**, 1169–1177 (1996).
- Seoh, S. A., Sigg, D., Papazian, D. M. & Bezanilla, F. Voltage-sensing residues in the S2 and S4 segments of the Shaker K<sup>+</sup> channel. *Neuron* **16**, 1159–1167 (1996).
- Tempel, B. L., Papazian, D. M., Schwarz, T. L., Jan, Y. N. & Jan, L. Y. Sequence of a probable potassium channel component encoded at Shaker locus of *Drosophila*. *Science* **237**, 770–775 (1987).
- Soler-Llavina, G. J., Chang, T. H. & Swartz, K. J. Functional interactions at the interface between voltage-sensing and pore domains in the Shaker K(v) channel. *Neuron* **52**, 623–634 (2006).
- Swartz, K. J. & MacKinnon, R. Hanatoxin modifies the gating of a voltage-dependent K<sup>+</sup> channel through multiple binding sites. *Neuron* **18**, 665–673 (1997).
- Swartz, K. J. & MacKinnon, R. Mapping the receptor site for hanatoxin, a gating modifier of voltage-dependent K<sup>+</sup> channels. *Neuron* **18**, 675–682 (1997).
- Li-Smerin, Y. & Swartz, K. J. Localization and molecular determinants of the hanatoxin receptors on the voltage-sensing domain of a K<sup>+</sup> channel. *J. Gen. Physiol.* **115**, 673–684 (2000).
- Li-Smerin, Y. & Swartz, K. J. Helical structure of the COOH terminus of S3 and its contribution to the gating modifier toxin receptor in voltage-gated ion channels. *J. Gen. Physiol.* **117**, 205–218 (2001).
- Lee, H. C., Wang, J. M. & Swartz, K. J. Interaction between extracellular hanatoxin and the resting conformation of the voltage-sensor paddle in Kv channels. *Neuron* **40**, 527–536 (2003).
- Phillips, L. R. *et al.* Voltage-sensor activation with a tarantula toxin as cargo. *Nature* **436**, 857–860 (2005).
- Swartz, K. J. Tarantula toxins interacting with voltage sensors in potassium channels. *Toxicon* **49**, 213–230 (2007).
- Ruta, V. & MacKinnon, R. Localization of the voltage-sensor toxin receptor on KvAP. *Biochemistry* **43**, 10071–10079 (2004).
- Jung, H. J. *et al.* Solution structure and lipid membrane partitioning of VSTx1, an inhibitor of the KvAP potassium channel. *Biochemistry* **44**, 6015–6023 (2005).
- Milescu, M. *et al.* Tarantula toxins interact with voltage sensors within lipid membranes. *J. Gen. Physiol.* **130**, 497–511 (2007).
- Lee, A. G. Lipid-protein interactions in biological membranes: a structural perspective. *Biochim. Biophys. Acta* **1612**, 1–40 (2003).
- Lampe, R. A. *et al.* Isolation and pharmacological characterization of omega-gammatoxin SIA, a novel peptide inhibitor of neuronal voltage-sensitive calcium channel responses. *Mol. Pharmacol.* **44**, 451–460 (1993).
- Lee, C. W. *et al.* Solution structure and functional characterization of SGTX1, a modifier of Kv2.1 channel gating. *Biochemistry* **43**, 890–897 (2004).
- Herrington, J. *et al.* Blockers of the delayed-rectifier potassium current in pancreatic  $\beta$ -cells enhance glucose-dependent insulin secretion. *Diabetes* **55**, 1034–1042 (2006).
- Papazian, D. M. *et al.* Electrostatic interactions of S4 voltage sensor in Shaker K<sup>+</sup> channel. *Neuron* **14**, 1293–1301 (1995).
- Long, S. B., Tao, X., Campbell, E. B. & MacKinnon, R. Atomic structure of a voltage-dependent K<sup>+</sup> channel in a lipid membrane-like environment. *Nature* doi:10.1038/nature06265 (this issue).
- Siemens, J. *et al.* Spider toxins activate the capsaicin receptor to produce inflammatory pain. *Nature* **444**, 208–212 (2006).
- Swartz, K. J. & MacKinnon, R. An inhibitor of the Kv2.1 potassium channel isolated from the venom of a Chilean tarantula. *Neuron* **15**, 941–949 (1995).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank F. Fontaine, M. Mayer, J. Mindell, S. Ramsey, S. Silberberg and members of the Swartz laboratory for discussions, and the NINDS DNA sequencing facility for DNA sequencing. We thank T. Kitaguchi for cloning KvAP and Y. Okamura for providing Ci-VSP complementary DNA. This work was supported by the Intramural Research Program of the NINDS, NIH. A.A.A. was partially supported by the NIH Undergraduate Scholarship Program.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to K.J.S. ([swartzk@ninds.nih.gov](mailto:swartzk@ninds.nih.gov)).

## METHODS

**Channel and chimaera constructs.** Chimaeras and point mutations were generated using sequential polymerase chain reaction (PCR) with KvAP<sup>24</sup>, Ci-VSP<sup>5</sup>, Hv1 (ref. 23), Kv2.1Δ7 (refs 3 and 25) or Shaker H4 (ref. 29) as templates. The Kv2.1Δ7 construct contains seven point mutations in the outer vestibule<sup>3</sup>, rendering the channel sensitive to agitoxin-2, a pore-blocking toxin from scorpion venom<sup>49</sup>, and the Shaker construct contains a deletion of residues 6 to 46 to remove fast inactivation<sup>50</sup>. KvAP was amplified from *A. pernix* genomic DNA (ATCC Inc.) and human Hv1 was amplified from clone NIH-MGC-92 (ID 5577070, Invitrogen). The DNA sequence of all constructs was confirmed by automated DNA sequencing, and cRNA was synthesized using T7 polymerase after linearizing DNA with appropriate restriction enzymes.

**Two-electrode voltage-clamp recording from *Xenopus* oocytes.** Channel constructs were expressed in *Xenopus* oocytes<sup>35</sup> and studied using two-electrode voltage-clamp recording techniques (OC-725C, Warner Instruments) with a 200-μl recording chamber. Data were filtered at 2 kHz and digitized at 10 kHz. Microelectrode resistances were 0.1–1 MΩ when filled with 3 M KCl. All experiments were performed at room temperature (~22 °C). One technical difficulty in studying the toxin sensitivity of chimaeras containing the entire KvAP paddle in Kv2.1 (C7[S3–S4]AP) is that these constructs display pronounced sensitivity to silver released from bath ground wires, resulting in pronounced inhibition when the flow of solution around the oocyte is stopped (for conserving the quantities of toxins used in experiments). We therefore limited our study of toxin–channel interactions to the chimaera containing the KvAP paddle in Shaker (C\*[S3–S4]AP), or chimaeras containing parts of the KvAP paddle in Kv2.1, constructs for which this technical problem is not pronounced. Leak and background conductances, identified by blocking the channel with agitoxin-2 (ref. 49), have been subtracted for all of the Kv channel currents shown<sup>35</sup>.

**Patch recording from HEK cells.** Macroscopic proton currents were recorded in whole-cell mode using a patch-clamp amplifier (Axopatch 200B). Data were filtered at 1 kHz or 2 kHz (8-pole Bessel filter) and digitized at 20 kHz. Patch pipette resistance when filled with the recording solution was 1–2 MΩ. When activating the Hv1 channel with voltage steps that are long enough to reach steady state (~2–3 s), the resulting proton currents were somewhat unstable when examined with repeated pulses to the same voltage, a phenomenon that we attribute to proton depletion near the membrane. To circumvent this problem we limited the pulse duration to 1.5 s and pulse frequency to 0.05 Hz, and studied cells with moderate expression levels—manipulations that result in relatively stable and reproducible proton currents. All experiments were performed at room temperature (~22 °C).

**Analysis of channel activity and toxin–channel interactions.** Voltage–activation relationships were obtained by measuring tail currents or steady-state currents (and calculating conductance), and a single Boltzmann function was fitted to the data according to:

$$I/I_{\max} = (1 + e^{-zF(V - V_{1/2})/RT})^{-1}$$

where  $I/I_{\max}$  is the normalized tail-current amplitude,  $z$  is the equivalent charge,  $V_{1/2}$  is the half-activation voltage,  $F$  is Faraday's constant,  $R$  is the gas constant and  $T$  is temperature in Kelvin.

Occupancy of closed or resting channels by tarantula toxins was examined using negative holding voltages where open probability was very low, and the fraction of unbound channels ( $F_u$ ) was estimated using depolarizations that are too weak to open toxin-bound channels, as described previously<sup>3,31–37,43</sup> (Supplementary Fig. 3). Because toxin-bound channels close or deactivate more rapidly than unbound channels, we examined the kinetics of deactivation using tail currents to confirm that toxin-bound channels did not contribute to the currents measured with weak depolarizations. After addition of the toxin to the recording chamber, the equilibration between the toxin and the channel was monitored using weak depolarizations elicited at 4–20-s intervals. For all channels, we recorded voltage–activation relationships (typically from tail currents) in the absence and presence of different concentrations of toxin. The ratio of currents ( $I/I_0$ ) recorded in the presence ( $I$ ) and absence ( $I_0$ ) of toxin was calculated for various strength depolarizations, typically –70 mV to +90 mV. The value of  $I/I_0$  measured in the plateau phase at voltages where toxin-bound channels do not open<sup>3,31–37,43</sup> was taken as  $F_u$  (see Supplementary Fig. 3 for an example). The apparent equilibrium dissociation constant ( $K_d$ ) was calculated assuming four independent toxin-binding sites per channel, with single occupancy being sufficient to inhibit opening in response to weak depolarizations.

$$K_d = ((1/(1 - F_u^{1/4}) - 1)[\text{Toxin}]$$

For all chimaeras and mutants, voltage protocols were adjusted appropriately so that the plateau phase in the  $I/I_0$ –voltage relationship was well defined (see Supplementary Fig. 3). Example traces showing the inhibitory activity of tarantula toxins were taken for relatively weak depolarizations within the plateau phase for that particular channel construct.

49. Garcia, M. L., Garcia-Calvo, M., Hidalgo, P., Lee, A. & MacKinnon, R. Purification and characterization of three inhibitors of voltage-dependent K<sup>+</sup> channels from *Leiurus quinquestriatus* var. *hebraeus* venom. *Biochemistry* **33**, 6834–6839 (1994).
50. Hoshi, T., Zagotta, W. N. & Aldrich, R. W. Biophysical and molecular mechanisms of Shaker potassium channel inactivation. *Science* **250**, 533–538 (1990).

## ARTICLES

# Atomic structure of a voltage-dependent K<sup>+</sup> channel in a lipid membrane-like environment

Stephen B. Long<sup>1†</sup>, Xiao Tao<sup>1</sup>, Ernest B. Campbell<sup>1</sup> & Roderick MacKinnon<sup>1</sup>

**Voltage-dependent K<sup>+</sup> (Kv) channels repolarize the action potential in neurons and muscle. This type of channel is gated directly by membrane voltage through protein domains known as voltage sensors, which are molecular voltmeters that read the membrane voltage and regulate the pore. Here we describe the structure of a chimaeric voltage-dependent K<sup>+</sup> channel, which we call the 'paddle-chimaera channel', in which the voltage-sensor paddle has been transferred from Kv2.1 to Kv1.2. Crystallized in complex with lipids, the complete structure at 2.4 ångström resolution reveals the pore and voltage sensors embedded in a membrane-like arrangement of lipid molecules. The detailed structure, which can be compared directly to a large body of functional data, explains charge stabilization within the membrane and suggests a mechanism for voltage-sensor movements and pore gating.**

Whether found on voltage-dependent ion channels<sup>1</sup>, voltage-dependent proton channels<sup>2,3</sup> or voltage-regulated enzymes<sup>4</sup>, amino acid sequence conservation indicates that the basic structural unit of the voltage sensor is conserved. The voltage sensor is a membrane-protein domain with charged amino acids embedded within the membrane electric field. It undergoes conformational change associated with displacement of charged amino acids, detectable through electrical measurements as gating charge<sup>5</sup>. The voltage sensor is therefore an electromechanical coupling device in which the membrane voltage biases the conformation of the domain, which in turn is coupled to the pore of an ion channel or to the active site of an enzyme.

X-ray crystallographic studies of Kv channels have described the voltage-sensor architecture, but many fundamental questions remain unanswered<sup>6–8</sup>. These questions concern the distance over which the gating charges move, and the physical and chemical principles by which the charges are stabilized within the membrane. Ultimately, these questions must be addressed using atomic structures in a membrane setting. To this end, our laboratory has explored three-dimensional protein crystals containing lipid–detergent mixtures to achieve membrane-protein structures in a membrane-like environment.

## Structure of the paddle-chimaera channel

We have determined the structure of a modified rat Kv1.2 K<sup>+</sup> channel in which the voltage-sensor paddle, a helix–turn–helix structural motif comprising the S3b and S4 helices, has been replaced by the voltage-sensor paddle from the rat Kv2.1 K<sup>+</sup> channel (Fig. 1a)<sup>9,10</sup>. Construction of the paddle-chimaera channel was based on the demonstration that the voltage-sensor paddle is transferable among voltage sensors of different origin<sup>11</sup>. The crystals, grown in a mixture of detergents and phospholipids, are unrelated to the Kv1.2 crystal and are formed with unique protein–protein contacts<sup>7</sup>. Diffraction data were collected to 2.4 ångström (Å), and phases were determined by molecular replacement using the Kv1.2 crystal structure as a search model. Electron density corresponding to the voltage-sensor paddle is shown (Fig. 1b). The maps were of high quality throughout, except for a segment connecting the cytoplasmic T1 domain to the first transmembrane region. This segment, built as a polyglycine helix

in the Kv1.2 structure, is probably extended over part of its length (Supplementary Fig. 1). The final model, refined to an  $R_{\text{free}}$  of 0.24, includes all amino acids within the transmembrane region, 16 total or partial lipid molecules per subunit, an NADP<sup>+</sup> cofactor and 193 water molecules (molecule 1, Methods).

An  $\alpha$ -carbon trace of the transmembrane region is shown, with the voltage-sensor paddles coloured red and a segment of the S1–S2 connection coloured pale blue (Fig. 1c). The voltage-sensor paddle is tilted away from the central axis of the voltage sensor, towards the lipid membrane, such that it makes minimal contact with the remainder of the channel. One entire face of the voltage-sensor paddle is exposed to the lipid membrane. The region of minimal contact corresponds to the largest transferable segment of the paddle defined in ref. 11. The S1–S2 loop segment (pale blue) contains a short  $\alpha$  helix that rests between two halves of the voltage sensor (that is, the voltage-sensor paddle defining one half and the S1 and S2 helices defining the other half). The paddle-chimaera channel produces voltage-dependent K<sup>+</sup> currents in lipid membranes that are very similar to those in the Kv1.2 channel (Fig. 1d, chimaera). Removal of the pale blue segment within the S1–S2 loop (Fig. 1c) has little effect on voltage-dependent opening (Fig. 1d,  $\Delta$ S1–S2 loop).

## Conservation of structure in K<sup>+</sup> channels

The Shaker K<sup>+</sup> channel and other closely related eukaryotic Kv channels have been the subjects of K<sup>+</sup> conduction and selectivity studies, whereas the *Streptomyces lividans* K<sup>+</sup> channel (KcsA) has provided the most detailed structural description of a K<sup>+</sup> channel<sup>12–15</sup>. Eukaryotic Shaker-like Kv channels and KcsA have almost identical K<sup>+</sup> selectivity filters (Fig. 2a). The root mean square deviation for all atoms in the filter (sequence TVGYG) is 0.25 Å. This extent of deviation probably corresponds to the limit of accuracy with which the crystallographic data define the atomic positions at 2.4 Å resolution<sup>16</sup>. At 2.4 Å, the paddle-chimaera channel shows that the selectivity filter of Kv channels is structurally identical to inward rectifier K<sup>+</sup> channels and KcsA (conductive conformation) to within a few tenths of an ångström<sup>15,16</sup>.

Figure 2b shows a superposition of transmembrane regions for the paddle chimaera and Kv1.2. The voltage-sensor paddles of Kv1.2 and

<sup>1</sup>Howard Hughes Medical Institute, Laboratory of Molecular Neurobiology and Biophysics, Rockefeller University, 1230 York Avenue, New York, New York 10065, USA. <sup>†</sup>Present address: Structural Biology Program, Memorial Sloan-Kettering Cancer Center, Box 414, 1275 York Avenue, New York, New York 10065, USA.



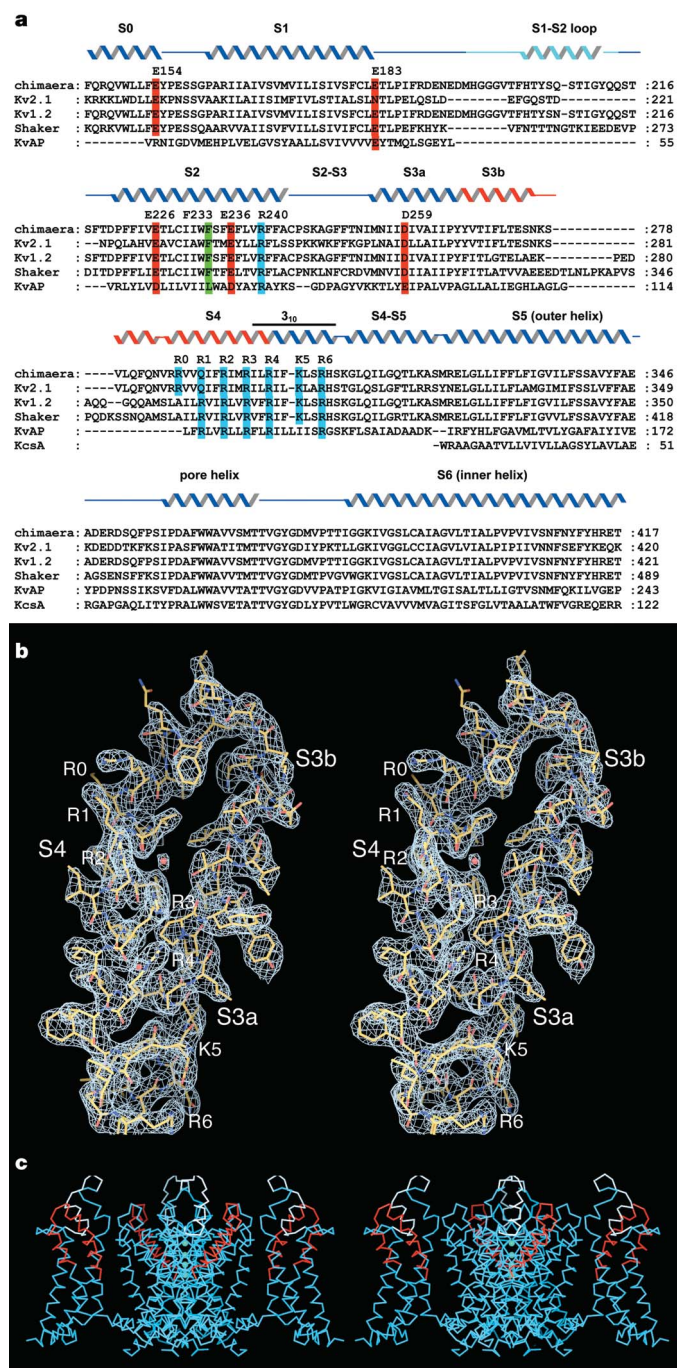
the paddle chimaera differ by 27 amino acids over the 32 amino acids substituted (Fig. 1a), but nevertheless their overall conformations are the same. In particular, the tilted position of the voltage-sensor paddle away from the central axis of the voltage sensor is very similar in the two structures. These two channels form different crystal lattices with unique protein contacts. Because the voltage sensor adopts the

same conformation in the settings of different crystal contacts, it is unlikely that these contacts dictate the conformation of the voltage sensor.

Figure 2c shows a superposition of the voltage sensor from the paddle-chimaera channel and the isolated voltage sensor of KvAP, a prokaryotic Kv channel<sup>6</sup>. For visual clarity, the S1–S2 loop of the paddle chimaera is not shown. The structures have been aligned to minimize the deviation between the S1 and S2 helices of the two channels. Specific positively and negatively charged amino acids almost superimpose, indicating that these voltage sensors are closely related members of the voltage-sensor structural family. However, there are informative structural differences between them. The sharp kink at a proline residue, which demarcates S3a and S3b in the paddle-chimaera channel, is replaced by a complete break and partial unwinding of the S3 helix in KvAP. The S2–S3 turn, also known as the membrane interface anchor<sup>6,8</sup>, is lifted 'higher' in KvAP, and the voltage-sensor paddles have different degrees of rotation. The paddle units themselves, however, can be superimposed as shown (Fig. 2d). The structural similarities between the paddle-chimaera and KvAP voltage sensors—in particular the good superposition of charged amino acids—imply that the voltage sensors are in related conformational states. The structural differences are compatible with the idea that the voltage-sensor paddle is mobile with respect to the S1 and S2 helices<sup>8,17,18</sup>.

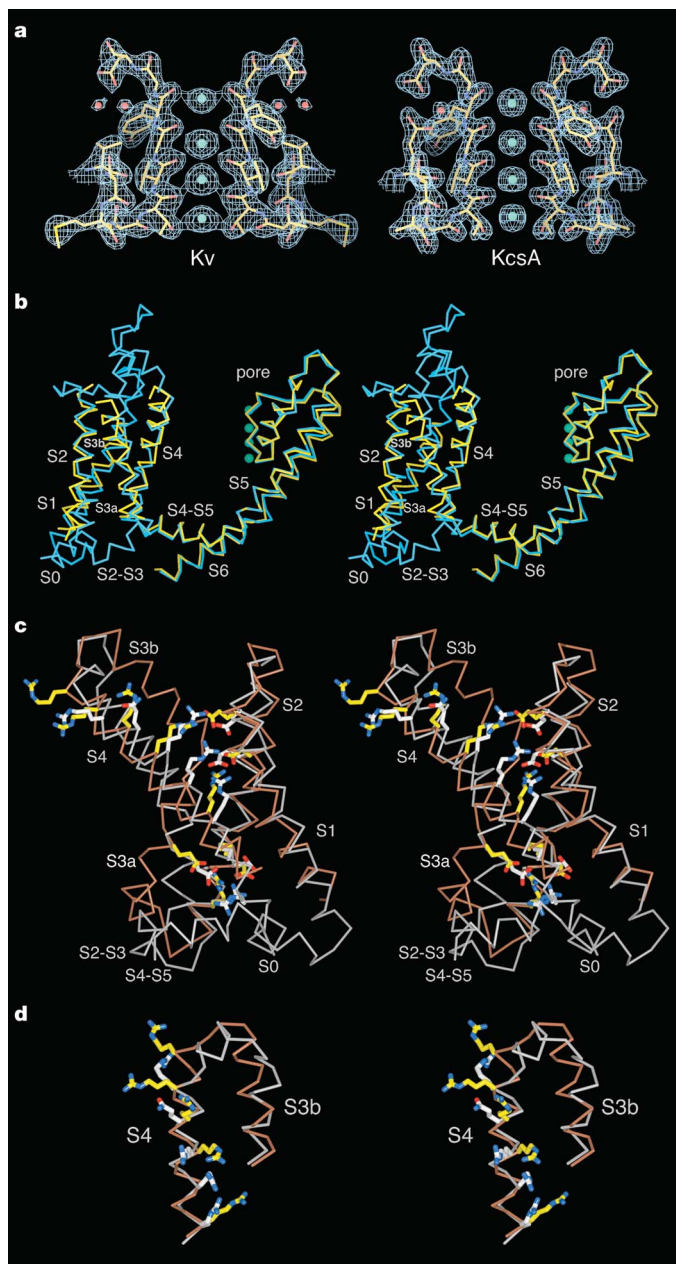
### Lipid interactions

Biochemical purification and crystallization of the Kv1.2 K<sup>+</sup> channel required a mixture of phospholipids and detergents<sup>7</sup>. The paddle-chimaera structure provides an explanation for the lipid requirement. Multiple complete and/or partial lipid molecules surround the channel in the crystal lattice (Fig. 3a, b and Supplementary Fig. 2). The extended alkyl chains of partial lipid molecules are distinguishable from detergent molecules because none of the detergents used in the preparation contained extended alkyl chains. Several aspects of the lipid molecules in the crystal are worth comment. First, there is a bilayer-like arrangement near the protein surface, with the outer leaflet particularly well formed (Fig. 3a). This observation suggests that the mixed micelle is a better membrane mimic than a pure detergent micelle. Second, lipid molecules are most dense in the concave hemi-circles between the voltage sensors (Fig. 3b). The hemi-circles reflect the very complex shape of a Kv channel, in which voltage-sensor domains protrude like appendages mostly surrounded by lipid membrane. By filling in space, the lipid molecules are an integral part of this membrane-protein structure and probably confer stability to it. Third, the lipid head groups reside at different 'membrane depths' with respect to the channel (Fig. 3a). This may, in part, reflect the fact that the bilayer-like micelle is not a true membrane. However, it may also reflect the fact that a membrane protein can perturb the positions of nearby lipid molecules away from the



**Figure 1 | Paddle-chimaera channel.** **a**, Sequence alignment of the paddle-chimaera channel, rat Kv2.1 (GI, 1235594), Shaker Kv (GI, 13432103), KvAP (GI, 5104624) and KcsA (GI, 61226909) transmembrane regions. Secondary structure elements are indicated above the sequences: pale blue, the region removed in the ΔS1–S2 loop construct; red, the S3b–S4 region of Kv2.1 inserted into the Kv1.2 channel (dark blue). Residues discussed in the text are highlighted red (negatively charged), blue (positively charged) and green (phenylalanine 233). **b**, Stereo representation of electron density (wire mesh) for the S3–S4 region (2F<sub>o</sub>–F<sub>c</sub> calculated from 50–2.4 Å using phases from the final model and contoured at 1.2σ). The structure of the protein (sticks) and two water molecules (red spheres, adjacent to R2 and 'behind' R4) are shown. **c**, Overall structure of the transmembrane region (α-carbon trace in stereo, coloured as in **a**). **d**, Channel activation. Currents are shown of wild-type Kv1.2 (left), the paddle-chimaera channel (middle), and the paddle-chimaera channel ΔS1–S2 loop (right), without subtracting leak and capacitive currents. Voltage pulses, –110 mV to 110 mV; ΔV = 10 mV; holding potential, –110 mV, stepping back to –90 mV.

more even plane of molecules normally associated with a pure lipid bilayer. The unusually positioned lipid molecules make specific favourable interactions with the ion channel, and probably influence its structure and function<sup>19</sup>. An example of a displaced phospholipid molecule is shown wedged between the voltage sensor and the S4–S5 linker, a region of the channel that couples voltage-sensor motions to pore gating (Fig. 3c)<sup>18,20,21</sup>.

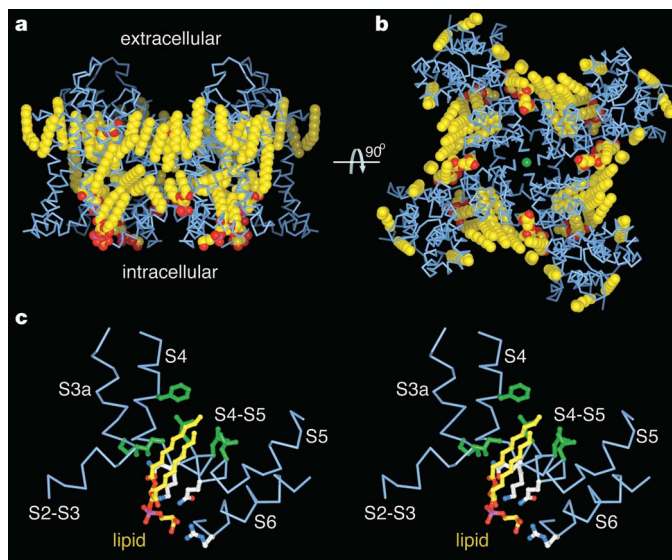


**Figure 2 | Comparison of the paddle-chimaera channel with other K<sup>+</sup> channel structures.** **a**, K<sup>+</sup> selectivity filter of the paddle-chimaera channel (Kv, left) compared with that of KcsA (PDB accession number, 1K4C, right). For clarity, two of the four subunits are shown (yellow sticks). Electron density (wire mesh,  $2F_o - F_c$ ,  $2.5\sigma$  contour), K<sup>+</sup> ions (green spheres) and water molecules (red spheres) are shown. **b**, Comparison with the Kv1.2 structure (**b**, **c** and **d** are shown in stereo). The  $\alpha$ -carbon trace of Kv1.2 (yellow, PDB accession number, 2A79) and paddle chimaera structures (cyan) are superimposed. **c**, Superposition of the paddle chimaera voltage sensor (grey  $\alpha$ -carbon trace, with the S1–S2 connecting loop removed for visual clarity), and the isolated voltage sensor of KvAP<sup>6</sup> (brown  $\alpha$ -carbon trace, PDB accession number, 1ORS). Positively and negatively charged amino acids are shown as sticks (KvAP, yellow; paddle chimaera, white). **d**, Superposition of the S3b–S4 voltage sensor paddles from KvAP and the chimaera channel (coloured as in **c**).

### Detailed structure of the voltage sensor

Figure 4a shows a voltage sensor and an S4–S5 linker helix (white  $\alpha$ -carbon trace) relative to the pore (cyan  $\alpha$ -carbon trace). The gate of the ion-conduction pore, which is formed by the crossing of four S6 inner helices (one from each subunit)<sup>22</sup>, adopts an open conformation in the paddle-chimaera channel. The S4–S5 linker helix engages the S6 helix and is angled in a manner consistent with the channel's open conformation, as was observed in the Kv1.2 structure<sup>18</sup>. The S1 helix makes specific contacts with the pore helix and the S5 helix (green side chains). Hanatoxin-interacting residues are located on the lipid-exposed face of the voltage-sensor paddle (red side chains)<sup>23</sup>. Figure 4b shows a voltage sensor and an S4–S5 linker helix viewed from the pore, with the yellow side chains from Fig. 4a and water molecules coloured according to atom type. The tilted position of the voltage-sensor paddle (S3b–S4) away from the S1 and S2 helices creates an external aqueous cleft in the voltage sensor that penetrates approximately 10 Å below the level of the membrane surface.

Negatively charged amino acids in the voltage sensor fall into two clusters: an external negative cluster consisting of glutamate 183 (S1 helix) and glutamate 226 (S2 helix), and an internal negative cluster consisting of glutamate 154 (S0 helix), glutamate 236 (S2 helix) and aspartate 259 (S3a helix) (Fig. 4b). These negative clusters are separated by approximately 15 Å (distance between nearest charges: glutamate 226 and glutamate 236). The external negative cluster is located in the external aqueous cleft. The internal negative cluster, in particular amino acids glutamate 236 and aspartate 259, are part of a buried network of charged amino acids, not exposed to aqueous solution, located approximately 10 Å from the intracellular membrane surface. Phenylalanine 233, located near the midpoint of the membrane, separates the external and internal negative clusters (Fig. 4b, green). Among Kv channels, the amino acids forming the negative clusters are very conserved as glutamate or aspartate. Phenylalanine 233 is the single most conserved amino acid in Kv



**Figure 3 | Lipids in the structure.** **a**, Side view of the transmembrane region showing lipid molecules (CPK representation, coloured according to atom type: yellow, carbon; red, oxygen; magenta, phosphorous) and the channel (blue  $\alpha$ -carbon trace). **b**, View from the extracellular side of the membrane with potassium shown as a green sphere. The lipid observed at the subunit–subunit interface in the structure of KcsA<sup>15</sup> (PDB accession number, 1K4C) is also observed in the paddle-chimaera channel (prominent red headgroup, nearest the centre). **c**, Detailed interactions of a lipid (yellow and red sticks) bound between the voltage sensor and the S4–S5 linker helix, in stereo. Amino acids that interact with the lipid are shown as sticks and are coloured according to chemical nature (green, hydrophobic residues; white, hydrophilic residues with nitrogen (blue) and oxygen (red) atoms).



channels outside the selectivity filter (Fig. 1a). KvAP is a rare example in which leucine is present in place of phenylalanine.

Positively charged amino acids on S4 are labelled in Fig. 4b according to an alignment with the Shaker  $K^+$  channel (Fig. 1a). Kv2.1 has a glutamine residue (glutamine 290 in the paddle chimera) corresponding to the first S4 arginine in Shaker (R1) and an additional preceding arginine labelled R0. Positively charged amino acids K5 and R6 form ionized hydrogen bonds with the internal negative cluster, whereas R3 and R4 form ionized hydrogen bonds with the external negative cluster. R0, R1 and R2 are positioned to interact favourably with the lipid phosphodiester layer, a mixed lipid–water environment and the water environment of the external aqueous cleft, respectively.

The side chains of R0, R1, R2, R3 and R4 are all effectively in or close to the extracellular solution, where we expect to find them in the open conformation of the voltage sensor. Four aspects of the voltage-sensor structure account for the extracellular disposition of these amino acids. First, the tilted paddle creates the external cleft into which the extracellular solution penetrates. Second, the side chains of R3 and R4 are drawn towards the surface by electrostatic attraction to the external negative cluster. Third, the S4 helix adopts a  $3_{10}$ -helix hydrogen-bonding pattern 'below' position 296 (R3)<sup>24</sup>. The  $3_{10}$  helix is associated with a translation of 2.0 Å per amino acid, in contrast to 1.5 Å for an  $\alpha$  helix<sup>25</sup>. Consequently, over a 10-amino-acid segment (position 297 to 306) the S4 helix is 5 Å longer than it would be if it were entirely  $\alpha$ -helical. In effect, the 'bottom' of the S4 helix is stretched towards the extracellular solution, allowing the voltage-sensor paddle to reside closer to the extracellular solution. Fourth, the short helix in the S2–S3 turn leading into S3a and the S4–S5 linker helix leading out of S4 are both angled in a manner to suggest that the

entire S3 plus S4 'half' of the voltage sensor is displaced towards the extracellular solution relative to the S1 plus S2 'half' (Fig. 4a, b).

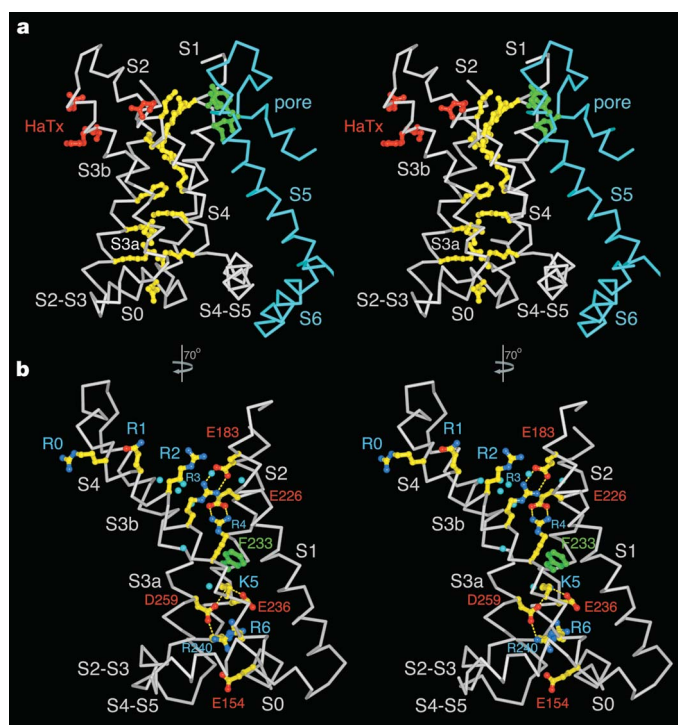
### Inferring motion from functional data

For the Shaker and Kv2.1  $K^+$  channels, opening is coupled to displacement of 3 or 4 charge units per voltage sensor (12 to 16 charges per channel) across the membrane-voltage difference<sup>26–29</sup>. The roughly equal total gating charge measured for Shaker and Kv2.1 suggests that R0 in Kv2.1 might compensate for the absence of a charge at R1 (Figs 1a and 4b)<sup>30</sup>. In Shaker, R1, R2, R3 and R4 each contribute approximately a single charge and K5 contributes approximately half a charge<sup>28,29</sup>. The crystal structure of the open conformation is consistent with these electrical measurements if all five positions reach an internal location in the closed conformation: the first four would thus move from outside to inside and the fifth would move from its midway point (about half way between the base of the external cleft and the internal solution) to inside (Fig. 4b). 'Outside' and 'inside' in these measurements refer to regions of electrical isopotential with bulk solution, which are a function of the detailed geometry of the voltage sensor and the surrounding lipid molecules.

Gating-associated-protein motions have been assessed using methanethiosulfonate (MTS) reactivity of site-directed cysteines in the Shaker  $K^+$  channel (Fig. 5a, b)<sup>31</sup>. Red, yellow and blue spheres correspond to positions accessible from the outside, not accessible from either side, and accessible from the inside, respectively. In the open conformation (Fig. 5a), accessibility is compatible with the crystal structure: yellow spheres (inaccessible) are distributed across a buried 17 Å stretch extending from the base of the external cleft to the internal solution. This entire stretch becomes exposed to internal MTS reagents (blue spheres) in the closed conformation (Fig. 5b), implying that closure is associated with a large translation of S4 towards the internal solution and/or a wide opening of the voltage sensor to the inside (enough to permit entry of MTS reagents all the way to R3).

Histidine substitutions imply motion by their state-dependent ability to transfer protons across the voltage sensor (Fig. 5c)<sup>32,33</sup>. At R4, a histidine substitution causes a proton current through the voltage sensor in the open conformation. In the crystal structure (open conformation), the R4  $\alpha$ -carbon (magenta sphere) is at the level of phenylalanine 233 (green). A histidine side chain at this location is well positioned to flip above and below the phenylalanine side chain by rapid rotamer exchange, allowing protons to be transferred from the base of the external cleft (outside) to protonatable side chains below. The charged network below the phenylalanine forms a continuous hydrogen-bonding network through which protons could diffuse all the way to the intracellular solution. Histidine substitution at R1 (red sphere) also creates a proton current, but only when the voltage sensor adopts its closed conformation. A simple interpretation is that voltage-sensor closure moves the R1  $\alpha$ -carbon next to phenylalanine 233 (that is, to the position of the R4  $\alpha$ -carbon in the open conformation). Such a motion would correspond to an S4 translation of approximately 15 Å. Histidine substitution at R2 and R3 (orange spheres) creates a proton shuttle that has maximum throughput at 50% open probability, implying that these positions are alternately exposed to one side or the other, depending on the conformation of the voltage sensor. This behaviour is expected if, indeed, R1 resides near phenylalanine 233 in the closed conformation.

In KvAP, motions have been inferred by accessibility experiments in which biotin, tethered at specific positions, is captured by avidin in solution (Fig. 5d)<sup>34</sup>. The main difference between this type of accessibility and MTS accessibility is that avidin is too large to enter clefts in the voltage sensor, so reactivity requires the tethered biotin to reach the membrane surface. Red, yellow and blue spheres correspond to positions accessible from the outside, not accessible from either side, and accessible from the inside, respectively, in either the



**Figure 4 | Details of the voltage sensor.** **a**, Stereo representation of a voltage sensor and the S4–S5 linker helix (white  $\alpha$ -carbon trace) in relation to the pore (cyan  $\alpha$ -carbon trace). The view is from the side (extracellular solution 'above' and intracellular solution 'below'). Select residues are shown as sticks and are coloured as indicated in the text. **b**, Voltage sensor and S4–S5 linker helix (white  $\alpha$ -carbon trace) viewed from the pore. Yellow side chains from **a** and water molecules are now coloured according to atom type (yellow, carbon; blue, nitrogen; red, oxygen; green, phenylalanine 233; cyan, water). Ionized hydrogen bonds between basic and acidic residues are indicated by dashed yellow lines.



opened or the closed conformation. Blue spheres on the S4 helix (near R4), which are near the extracellular surface in the open conformation, must move to the level of the other blue spheres near the intracellular surface when the channel closes. This would necessitate a translation of at least 15 Å.

These different electrical and biochemical measurements interpreted in light of the crystal structure are consistent, and point to an S4 displacement of about 15 Å and a large reorganization of the voltage sensor associated with gating.

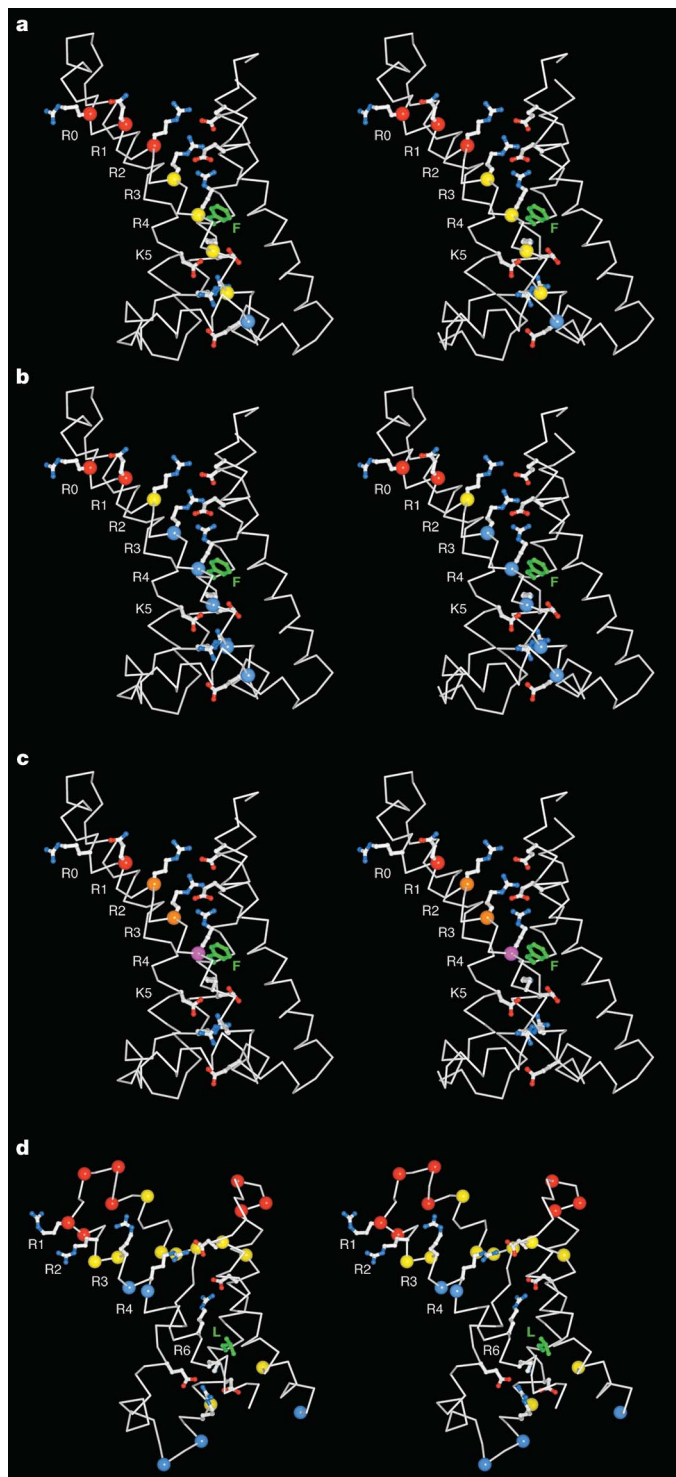
### Hypothesis for the voltage-sensing mechanism

Electrophysiological measurements show that the paddle-chimaera channel is open at 0 mV (Fig. 1d). In the crystal, which is at 0 mV and

provides a membrane-like environment, the channel also adopts an open conformation. Closing the voltage sensor requires a strong electric field associated with a negative (intracellular) membrane voltage to force the positive gating charges inward. Electrical measurements indicate that the charges get most of the way across the voltage difference during this transition<sup>28,29</sup>. Until the structure of a closed conformation is solved, the details of how this occurs will remain unknown; however, the pictures in Fig. 6 illustrate one way in which this could happen. In the depiction, the voltage-sensor paddle moves relative to the S1 and S2 helices. This movement translates the region of S4 centred on the gating charges ~15 Å towards the intracellular side, approximately half the membrane hydrophobic core thickness of a pure phosphatidylcholine bilayer (30 Å)<sup>35</sup>. Because of the shape of the voltage sensor and its effect on membrane thickness, 15 Å would be sufficient to transfer the charges from the electrical outside to the electrical inside.

This model is silent on many details, but further inspection of the crystal structure (Fig. 4b) stimulates several new ideas concerning voltage-sensor motion. First, the S2–S3 turn looks as if it could reorganize its structure, enabling movements within the voltage sensor (Fig. 4a, b). This turn is conserved in length and in certain amino acids (Fig. 1a), but is one of the least-studied regions in Kv channels. Second, the contact between S1 and the pore (Fig. 4a, green) is well positioned to hold the S1–S2 half of the voltage sensor static relative to the pore. Without this contact, motions of the voltage-sensor paddle relative to S1 and S2 might not transmit force as effectively to the S4–S5 linker helix. Ultimately, sensor closure must push down on the S4–S5 linker helix to constrict the pore until it is closed, as depicted (Fig. 6)<sup>18</sup>. Third, the 3<sub>10</sub>-helical intracellular half of S4, in addition to allowing the voltage-sensor paddle to reside closer to the extracellular solution in the open state (by ‘stretching’ S4), also directs arginine and lysine side chains that occur at every third position to the same helical face<sup>25</sup>. In the crystal structure, in which the channel is open, the 3<sub>10</sub> helix directs R4, K5 and R6 away from the membrane and into the sensor (Fig. 4b). It is possible that during closure, as the paddle moves inward, the zone of 3<sub>10</sub>-helical secondary structure shifts as a moving segment along S4 in a wave-like fashion. Such a ‘concertina effect’ would turn arginine residues away from the lipid membrane and towards the inside of the voltage sensor as they cross the centre of the hydrophobic core without necessitating rigid body rotation of the S4 helix.

The crystal structure also provides new ideas concerning charge stabilization during gating. Two negatively charged clusters near the external and internal aspects of the voltage sensor capture and stabilize gating charges on either side (Fig. 4b). But how are charges stabilized while crossing between these clusters? Perhaps the most intriguing aspect of the voltage sensor is a complete absence of charge near the membrane centre for a hydrophobic zone of roughly 10 Å between the negative clusters (Fig. 4b). This zone contains a phenylalanine residue (position 233 in the paddle-chimaera channel) in 355



**Figure 5 | Mapping of biochemical data on the voltage sensor.** In **a–c**, the voltage sensor and S4–S5 linker helix (white  $\alpha$ -carbon trace) are viewed from the pore (extracellular solution ‘above’ and intracellular solution ‘below’), in stereo. Phenylalanine 233 (F) is coloured green. Basic and acidic residues (sticks) are coloured according to atom type: white, carbon; red, oxygen; blue, nitrogen. **a, b**, MTS reactivity of site-directed cysteine residues introduced into Shaker on the S4 helix<sup>31</sup> mapped on the crystal structure. **a**, MTS accessibility in the open (depolarized) state. **b**, MTS accessibility in the closed (polarized) state. **c**, Histidine substitutions at four positions on the S4 helix<sup>32,33</sup> mapped on the structure (coloured spheres as described in the text). **d**, Biotin–avidin accessibility data mapped on the structure of the isolated voltage sensor of KvAP (PDB accession number, 1ORS)<sup>34</sup>. The view of voltage sensor (S1–S4, white  $\alpha$ -carbon trace) is the same as that for the paddle chimaera in **a–c** (superposition as in Fig. 2c). KvAP has a leucine (L, green side chain) at the position corresponding to phenylalanine 233 in the chimaera paddle. Coloured spheres indicate positions described in the text.

out of 360 Kv channels, and therefore we call it the 'phenylalanine gap' separating the external and internal negative clusters. Other than weak stabilization that might occur transiently through cation- $\pi$  interactions with the phenylalanine<sup>36</sup>, the phenylalanine gap seems energetically unfavourable for a positive charge. The arginine side chains must apparently flip between charge clusters on their way across the membrane. We speculate that the absence of polar groups in the central zone might prevent proton transfer across the membrane, which would lead to cell acidification, and it might also create an energy barrier for the arginines, which would favour a switch-like transition between closed and opened conformations.

## Discussion

The KvAP crystal structures and functional studies led to the hypothesis of a voltage-sensor paddle—a helix–turn–helix unit that moves at the protein–lipid interface<sup>6,17</sup>. It was proposed that acidic amino acids would help to stabilize positive charges on the paddle as it

moves within the membrane to control pore gating<sup>17</sup>. The Kv1.2 crystal structure demonstrated the lipid-surrounded, independent domain property of the voltage sensor, partial shielding of charges from the membrane, and the mechanical connection by which conformational changes in the sensor are transmitted to the pore through the S4–S5 linker helix<sup>7,18</sup>. Here, at high resolution and in a membrane-like environment, the paddle-chimaera channel reveals atomic interactions: arginine exposure to the membrane at the phosphodiester layer, charge stabilization by external and internal negative clusters inside the sensor, a phenylalanine gap across which the arginine residues must transit, and the possibility of a concertina effect in which an  $\alpha$ - to  $3_{10}$ -helical secondary structure transition could turn arginine residues away from the lipid membrane as paddle displacement moves them across the hydrophobic core.

## METHODS SUMMARY

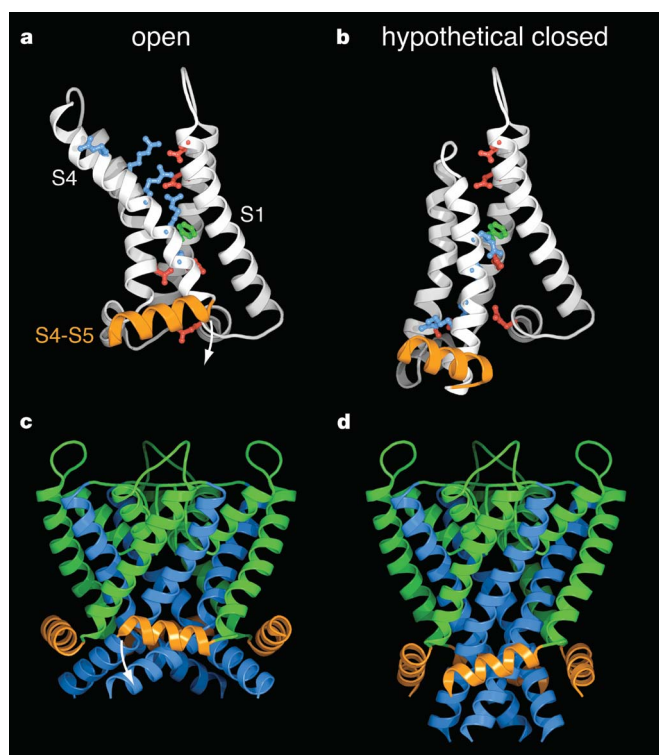
The paddle-chimaera channel with a His<sub>10</sub> tag at the amino terminus was co-expressed with the rat  $\beta$ 2.1 subunit in *Pichia pastoris* and purified using metal affinity chromatography and gel filtration. Protein crystals were grown from hanging drops at pH 8.5, and contain the detergents 6-cyclohexyl-1-hexyl- $\beta$ -D-maltoside (Cymal-6), 7-cyclohexyl-1-heptyl- $\beta$ -D-maltoside (Cymal-7) and CHAPS and the lipids POPC (1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine), POPE (1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine) and POPG (palmitoyl-2-oleoyl-*sn*-glycero-3-[phospho-rac-(1-glycerol)]). X-ray data (50–2.4 Å, space group *P*4<sub>2</sub>1<sub>2</sub>) were collected at beamline X29 (Brookhaven National Synchrotron Light Source (NSLS)). The structure was determined by molecular replacement using the structure of Kv1.2 in complex with its  $\beta$  subunit as a search model (PDB accession number, 2A79). There are two channel subunits and two  $\beta$  subunits in the asymmetric unit, giving rise to two tetrameric assemblies (identified as molecule 1 and molecule 2 in the text). The model was built using O<sup>37</sup> and was refined with CNS<sup>38</sup> to an  $R_{\text{free}}$  of 24.4%. Crystallographic data and refinement statistics are shown in the Supplementary Information.

The purified channels were reconstituted into POPE:POPG lipid vesicles by dialysis. Planar lipid bilayer experiments were performed as described previously<sup>39,40</sup>. Electrical measurements were carried out using the voltage-clamp method in whole-cell mode. Charybdotoxin (CTX) (1  $\mu$ M, ref. 41) was used to block channels in one orientation while channels in the other orientation were being recorded.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 8 August; accepted 17 September 2007.

- Noda, M. *et al.* Expression of functional sodium channels from cloned cDNA. *Nature* **322**, 826–828 (1986).
- Ramsey, I. S., Moran, M. M., Chong, J. A. & Clapham, D. E. A voltage-gated proton-selective channel lacking the pore domain. *Nature* **440**, 1213–1216 (2006).
- Sasaki, M., Takagi, M. & Okamura, Y. A voltage sensor-domain protein is a voltage-gated proton channel. *Science* **312**, 589–592 (2006).
- Murata, Y., Iwasaki, H., Iwasaki, M., Inaba, K. & Okamura, Y. Phosphoinositide phosphatase activity coupled to an intrinsic voltage sensor. *Nature* **435**, 1239–1243 (2005).
- Armstrong, C. M. & Bezanilla, F. Charge movement associated with the opening and closing of the activation gates of the Na<sup>+</sup> channels. *J. Gen. Physiol.* **63**, 533–552 (1974).
- Jiang, Y. *et al.* X-ray structure of a voltage-dependent K<sup>+</sup> channel. *Nature* **423**, 33–41 (2003).
- Long, S. B., Campbell, E. B. & MacKinnon, R. Crystal structure of a mammalian voltage-dependent Shaker family K<sup>+</sup> channel. *Science* **309**, 897–903 (2005).
- Lee, S. Y., Lee, A., Chen, J. & MacKinnon, R. Structure of the KvAP voltage-dependent K<sup>+</sup> channel and its dependence on the lipid membrane. *Proc. Natl Acad. Sci. USA* **102**, 15441–15446 (2005).
- Stuhmer, W. *et al.* Molecular basis of functional diversity of voltage-gated potassium channels in mammalian brain. *EMBO J.* **8**, 3235–3244 (1989).
- Frech, G. C., VanDongen, A. M., Schuster, G., Brown, A. M. & Joho, R. H. A novel potassium channel with delayed rectifier properties isolated from rat brain by expression cloning. *Nature* **340**, 642–645 (1989).
- Alabi, A. A., Bahamonde, M. I., Jung, H. J., Kim, J. I. & Swartz, K. J. Portability of paddle motif function and pharmacology in voltage sensors. *Nature* doi:10.1038/nature06266 (this issue).
- Kamb, A., Iverson, L. E. & Tanouye, M. A. Molecular characterization of Shaker, a *Drosophila* gene that encodes a potassium channel. *Cell* **50**, 405–413 (1987).
- Tempel, B. L., Papazian, D. M., Schwarz, T. L., Jan, L. Y. & Jan, Y. N. Sequence of a probable potassium channel component encoded at Shaker locus of *Drosophila*. *Science* **237**, 770–775 (1987).



**Figure 6 | Hypothetical mechanism of voltage-dependent gating.**

**a**, Representation of the voltage sensor and S4–S5 linker helix from the crystal structure (open conformation). Helices are drawn as ribbons. The view is from the pore, as in Fig. 5, with the extracellular solution 'above' and the intracellular solution 'below'. The gating charges (R1 to K5) are shown as blue sticks. Negatively charged residues in the external and internal clusters are red; the phenylalanine in the middle is green. The positively charged residues reach 'outward' towards the extracellular solution. **b**, Depiction of a hypothetical closed conformation of the voltage sensor. The S1 and S2 helices are hypothesized to maintain their position, whereas the S3–S4 paddle has moved inward. The positive charges on S4 now reach towards the intracellular solution, and are stabilized through interactions with the internal negative cluster. The  $\alpha$ -carbon position of R1 is adjacent to the phenylalanine, representing a displacement perpendicular to the plane of the membrane of approximately 15 Å relative to its location in the open structure (**a**). The inward displacement of the S4 helix pushes down on the N-terminal end of the S4–S5 linker helix, causing it to tilt towards the intracellular side and to close the pore. **c**, Depiction of the open conformation of the S4–S5 linker helices and pore from the crystal structure. The S4–S5 linker helices (orange) rest on the S6 helices (blue ribbons) near the intracellular side. **d**, A hypothetical model of the S4–S5 linker helices and pore in a closed conformation based on the crystal structure of a closed potassium channel pore (KcsA, PDB accession number, 1K4C)<sup>18</sup>.

14. Pongs, O. *et al.* Shaker encodes a family of putative potassium channel proteins in the nervous system of *Drosophila*. *EMBO J.* **7**, 1087–1096 (1988).
15. Zhou, Y., Morais-Cabral, J. H., Kaufman, A. & MacKinnon, R. Chemistry of ion coordination and hydration revealed by a K<sup>+</sup> channel–Fab complex at 2.0 Å resolution. *Nature* **414**, 43–48 (2001).
16. Nishida, M., Cadene, M., Chait, B. T. & MacKinnon, R. Crystal of a Kir3.1–prokaryotic Kir channel chimera. *EMBO J.* **26**, 4005–4015 (2007).
17. Jiang, Y., Ruta, V., Chen, J., Lee, A. & MacKinnon, R. The principle of gating charge movement in a voltage-dependent K<sup>+</sup> channel. *Nature* **423**, 42–48 (2003).
18. Long, S. B., Campbell, E. B. & MacKinnon, R. Voltage sensor of Kv1.2: structural basis of electromechanical coupling. *Science* **309**, 903–908 (2005).
19. Schmidt, D., Jiang, Q. X. & MacKinnon, R. Phospholipids and the origin of cationic gating charges in voltage sensors. *Nature* **444**, 775–779 (2006).
20. Lu, Z., Klem, A. M. & Ramu, Y. Coupling between voltage sensors and activation gate in voltage-gated K<sup>+</sup> channels. *J. Gen. Physiol.* **120**, 663–676 (2002).
21. Sukhareva, M., Hackos, D. H. & Swartz, K. J. Constitutive activation of the Shaker Kv channel. *J. Gen. Physiol.* **122**, 541–556 (2003).
22. Jiang, Y. *et al.* The open pore conformation of potassium channels. *Nature* **417**, 523–526 (2002).
23. Swartz, K. J. & MacKinnon, R. Mapping the receptor site for hanatoxin, a gating modifier of voltage-dependent K<sup>+</sup> channels. *Neuron* **18**, 675–682 (1997).
24. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).
25. Branden, C. & Tooze, J. *Introduction to Protein Structure* page 15 (Garland Publishing Inc., New York, 1999).
26. Zagotta, W. N., Hoshi, T., Dittman, J. & Aldrich, R. W. Shaker potassium channel gating. II: transitions in the activation pathway. *J. Gen. Physiol.* **103**, 279–319 (1994).
27. Schoppa, N. E., McCormack, K., Tanouye, M. A. & Sigworth, F. J. The size of gating charge in wild-type and mutant Shaker potassium channels. *Science* **255**, 1712–1715 (1992).
28. Aggarwal, S. K. & MacKinnon, R. Contribution of the S4 segment to gating charge in the Shaker K<sup>+</sup> channel. *Neuron* **16**, 1169–1177 (1996).
29. Seoh, S. A., Sigg, D., Papazian, D. M. & Bezanilla, F. Voltage-sensing residues in the S2 and S4 segments of the Shaker K<sup>+</sup> channel. *Neuron* **16**, 1159–1167 (1996).
30. Islas, L. D. & Sigworth, F. J. Voltage sensitivity and gating charge in Shaker and Shab family potassium channels. *J. Gen. Physiol.* **114**, 723–742 (1999).
31. Larsson, H. P., Baker, O. S., Dhillon, D. S. & Isacoff, E. Y. Transmembrane movement of the Shaker K<sup>+</sup> channel S4. *Neuron* **16**, 387–397 (1996).
32. Starace, D. M. & Bezanilla, F. A proton pore in a potassium channel voltage sensor reveals a focused electric field. *Nature* **427**, 548–553 (2004).
33. Starace, D. M. & Bezanilla, F. Histidine scanning mutagenesis of basic residues of the S4 segment of the shaker K<sup>+</sup> channel. *J. Gen. Physiol.* **117**, 469–490 (2001).
34. Ruta, V., Chen, J. & MacKinnon, R. Calibrated measurement of gating-charge arginine displacement in the KvAP voltage-dependent K<sup>+</sup> channel. *Cell* **123**, 463–475 (2005).
35. King, G. I. & White, S. H. Determining bilayer hydrocarbon thickness from neutron diffraction measurements using strip-function models. *Biophys. J.* **49**, 1047–1054 (1986).
36. Burley, S. K. & Petsko, G. A. Amino–aromatic interactions in proteins. *FEBS Lett.* **203**, 139–143 (1986).
37. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.* **A47**, 110–119 (1991).
38. Brunger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr.* **D54**, 905–921 (1998).
39. Ruta, V., Jiang, Y., Lee, A., Chen, J. & MacKinnon, R. Functional analysis of an archaeobacterial voltage-dependent K<sup>+</sup> channel. *Nature* **422**, 180–185 (2003).
40. Miller, C. *Ion Channel Reconstitution* (Plenum, New York, 1986).
41. Miller, C., Moczydlowski, E., Latorre, R. & Phillips, M. Charybdotoxin, a protein inhibitor of single Ca<sup>2+</sup>-activated K<sup>+</sup> channels from mammalian skeletal muscle. *Nature* **313**, 316–318 (1985).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank A. L. MacKinnon for CTX and discussions, S.-Y. Lee and T. Muir for discussions, K. Swartz for providing information enabling the chimera construction, and the staff at beamline X29, NSLS, Brookhaven National Laboratory. R.M. is an Investigator in the Howard Hughes Medical Institute. This work was supported by the NIH (R.M.).

**Author Information** The X-ray crystallographic coordinates and structure factor data have been deposited in the Protein Data Bank under accession number 2R9R. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to R.M. ([mackinn@mail.rockefeller.edu](mailto:mackinn@mail.rockefeller.edu)).



## METHODS

**Cloning, expression and purification.** A gene encoding the paddle-chimaera channel was constructed using standard DNA techniques as follows. The full-length rat Kv1.2 gene<sup>9</sup> (GI, 1235594) with an N-terminal His<sub>10</sub> tag and a thrombin protease cleavage site (MAH<sub>10</sub>GLVPRGSM(TX)<sub>n</sub>stop) was ligated into the multiple cloning site of the pPICZ-C vector (Invitrogen Life Technologies). Five amino acid substitutions were made to remove a glycosylation site and to reduce sensitivity to oxidation (N207Q, C31S, C32S, C434S and C482S). The region of the Kv1.2 gene encoding the S3b–S4 paddle was replaced with the corresponding region of rat Kv2.1 to yield the protein sequence VAIIIPYYVTIFLTESNKS<sup>SVLQFQNVRRVVQIFRIMRILRIFK</sup> (in which the Kv2.1 sequence is underlined). This construct encoding the paddle-chimaera channel was combined with a pPICZ-C vector containing the rat  $\beta$ 2.1 gene (residues 36–367) as described<sup>7</sup>. This vector, containing both the Kv1.2–Kv2.1 paddle-chimaera channel and the  $\beta$ 2.1 subunits, was used to transform yeast. Transformation of *P. pastoris*, expression, and cobalt affinity purification were performed as described<sup>7</sup>.

After elution from the cobalt affinity column, the protein was concentrated to approximately 20 mg ml<sup>-1</sup>, and was further purified on a Superdex-200 gel-filtration column. The buffer used for gel filtration consisted of: 150 mM KCl, 20 mM TRIS-HCl, pH 7.5, 3 mM Cymal-6 (anagrade, Anatrace), 3 mM Cymal-7 (anagrade, Anatrace), 1 mM EDTA, 2 mM Tris[2-carboxyethyl]phosphine (TCEP), 10 mM DTT and 0.1 mg ml<sup>-1</sup> lipid (3:1:1, POPC:POPE:POPG, obtained from Avanti). The mixture of detergents used (Cymal-6 and Cymal-7) improved the diffraction quality of the crystals. The gel-filtration fraction eluting between 11 ml and 12 ml showed Kv1.2 and  $\beta$ 2.1 proteins in correct ratio on Coomassie-stained SDS-PAGE. Protein was concentrated to 10 mg ml<sup>-1</sup> (using a Centricon-50, Millipore), mixed with crystallization solution (1  $\mu$ l protein plus 1  $\mu$ l solution), supplemented with 0.1  $\mu$ l CHAPS detergent additive (80 mM Hampton Research) and crystallized using the hanging-drop method over reservoirs containing 0.1 ml crystallization solution at 20 °C. The crystallization solution contained 28–36% PEG 400, 50 mM TRIS-HCl, pH 8.5. Crystals appeared within 5 days and were approximately 200  $\mu$ m in each dimension.

**$\Delta$ S1–S2 loop construct.** The paddle-chimaera channel construct was used to make a construct in which a portion of the loop connecting the S1 and S2 helices was removed (Fig. 1a, b). The amino acid residues 196–213 were replaced with three glycine residues to yield the protein sequence: DENEDGGGQSTSF. This protein was expressed and purified as described above, and was used for electrophysiology experiments.

**Structure determination.** Crystals were frozen in liquid nitrogen directly from their mother liquor. Diffraction data were collected to 2.4 Å at beamline X29 (Brookhaven NSLS), images were processed with DENZO, and intensities were merged with SCALEPACK<sup>42</sup>. Data were further processed using the CCP4 suite<sup>43</sup>. The crystals belong to the *P4*<sub>2</sub>12 space group. The structure was solved by molecular replacement using the program MOLREP<sup>44</sup>, with the 2.9 Å resolution structure of rat Kv1.2 in complex with rat  $\beta$ 2.1 subunit

(PDB accession number, 2A79) as a search model. There are two channel subunits and two  $\beta$  subunits in the asymmetric unit, which, with crystallographic symmetry, gave rise to two four-fold tetramers (molecules 1 and 2). The tetramerization (T1) and  $\beta$ -subunit regions of these two tetramers are identical within the error of the coordinates. The electron density for the voltage sensors of one of the tetramers (transmembrane chain B) was significantly better defined than the other and was used to direct model building. The model was built using O<sup>37</sup> and was refined with CNS<sup>38</sup> from 50–2.4 Å to an *R*<sub>free</sub> of 24.4%. The final model contains  $\beta$ 2.1 residues 36–361 and paddle-chimaera channel residues 32–417. All side chains are included in the model except channel residues 133–144 (the link between T1 and the transmembrane region), which were modelled as poly-glycine. Crystallographic data and refinement statistics are shown in Supplementary Information. Figures were made using PyMol (<http://www.pymol.org>).

**Reconstitution of Kv1.2 channels.** The purified channels were reconstituted into lipid vesicles using a published procedure<sup>45</sup> with modifications. In brief, POPE and POPG lipid mixture (15 mg ml<sup>-1</sup>:5 mg ml<sup>-1</sup>, respectively, in reconstitution buffer containing 10 mM HEPES, pH 7.0, 450 mM KCl and 2 mM DTT) were solubilized with 8% octyl- $\beta$ -D-maltopyranoside (Anatrace). The protein (2–10 mg ml<sup>-1</sup> in gel filtration buffer, above) was then mixed with an equal volume of the solubilized lipids to yield a final protein concentration of 1–5 mg ml<sup>-1</sup> and a lipid concentration of 10 mg ml<sup>-1</sup>. Detergent was removed by dialyzing against reconstitution buffer with 2 mM freshly added DTT at 4 °C for 4–6 days. The resulting protein–lipid reconstitution was flash-frozen in liquid N<sub>2</sub> and stored at –80 °C as small aliquots.

**Electrophysiological recordings.** Bilayer experiments were performed as described previously<sup>39</sup>. In brief, planar lipid bilayers of POPE:POPG (15 mg ml<sup>-1</sup>:5 mg ml<sup>-1</sup>) in decane were formed by painting over a 300- $\mu$ m hole in a polystyrene partition separating two aqueous chambers<sup>40</sup>. The chamber to which vesicles were added (*cis*) contained 4 ml of 150 mM KCl and 10 mM HEPES, pH 7.0; the opposite side (*trans*) contained 3 ml of 150 mM KCl and 10 mM HEPES (pH 7.0). Measurements were carried out using the voltage-clamp method in whole-cell mode, an Axopatch 200B amplifier, a Digidata 1322A analogue-to-digital converter and Axoclamp software (Axon Instruments) to control membrane voltage and record currents. CTX (1  $\mu$ M)<sup>41</sup> was used to block channels in one orientation while channels in the other orientation were being recorded.

42. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
43. Collaborative Computational Project. The CCP4 suite: programs for X-ray crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
44. Vagin, A. & Teplyakov, A. An approach to multi-copy search in molecular replacement. *Acta Crystallogr. D* **56**, 1622–1624 (2000).
45. Heginbotham, L., LeMasurier, M., Kolmakova-Partensky, L. & Miller, C. Single streptomyces lividans K<sup>+</sup> channels: functional asymmetries and sidedness of proton activation. *J. Gen. Physiol.* **114**, 551–560 (1999).

# Crystal structure of the human $\beta_2$ adrenergic G-protein-coupled receptor

Søren G. F. Rasmussen<sup>1\*</sup>, Hee-Jung Choi<sup>1,2\*</sup>, Daniel M. Rosenbaum<sup>1\*</sup>, Tong Sun Kobilka<sup>1</sup>, Foon Sun Thian<sup>1</sup>, Patricia C. Edwards<sup>3</sup>, Manfred Burghammer<sup>4</sup>, Venkata R. P. Ratnala<sup>1</sup>, Ruslan Sanishvili<sup>5</sup>, Robert F. Fischetti<sup>5</sup>, Gebhard F. X. Schertler<sup>3</sup>, William I. Weis<sup>1,2</sup> & Brian K. Kobilka<sup>1</sup>

**Structural analysis of G-protein-coupled receptors (GPCRs) for hormones and neurotransmitters has been hindered by their low natural abundance, inherent structural flexibility, and instability in detergent solutions. Here we report a structure of the human  $\beta_2$  adrenoceptor ( $\beta_2$ AR), which was crystallized in a lipid environment when bound to an inverse agonist and in complex with a Fab that binds to the third intracellular loop. Diffraction data were obtained by high-brilliance microcrystallography and the structure determined at 3.4 Å/3.7 Å resolution. The cytoplasmic ends of the  $\beta_2$ AR transmembrane segments and the connecting loops are well resolved, whereas the extracellular regions of the  $\beta_2$ AR are not seen. The  $\beta_2$ AR structure differs from rhodopsin in having weaker interactions between the cytoplasmic ends of transmembrane (TM)3 and TM6, involving the conserved E/DRY sequences. These differences may be responsible for the relatively high basal activity and structural instability of the  $\beta_2$ AR, and contribute to the challenges in obtaining diffraction-quality crystals of non-rhodopsin GPCRs.**

GPCRs represent the largest family of membrane proteins in the human genome. They are remarkably versatile signalling molecules that are responsible for the majority of transmembrane signal transduction in response to hormones and neurotransmitters. GPCRs share a common structural signature of seven membrane-spanning helices with an extracellular N terminus and an intracellular C terminus (Fig. 1). Our understanding of GPCR structure has been based largely on the crystal structures of the inactive state of rhodopsin<sup>1–7</sup>. Rhodopsin is better suited for structural studies than most other GPCRs because it is possible to obtain large quantities of functional protein from bovine retina. Rhodopsin is also a remarkably stable GPCR, retaining function under conditions that denature other GPCRs.

The  $\beta_2$ AR is a GPCR activated by adrenaline that plays important parts in cardiovascular and pulmonary physiology, and is one of the most extensively characterized members of this large family of membrane proteins<sup>8</sup>. The sites of interactions between agonists and the receptor have been characterized by mutagenesis studies<sup>9–12</sup>, and biophysical methods have been used to study the conformational changes associated with agonist binding and activation<sup>13–18</sup>. The  $\beta_2$ AR is efficiently expressed in Sf9 insect cells and can be purified to homogeneity using antibody and ligand affinity chromatography<sup>19</sup>. The  $\beta_2$ AR is biochemically pure following chromatography using an antibody resin that binds to an N-terminal Flag epitope; however, more than half of these purified receptor molecules are not functional. Affinity chromatography, an important early development in GPCR biochemistry<sup>20</sup>, is essential for isolating functional  $\beta_2$ AR protein. Purified  $\beta_2$ AR bound to an antagonist remains stable and soluble at concentrations up to 50 mg ml<sup>−1</sup> for up to a week at room temperature in the detergent dodecylmaltoside. However, the  $\beta_2$ AR is unstable in detergents used to obtain crystals of bovine rhodopsin. Extensive sparse matrix screening (over 2,000 conditions at 4 °C and 20 °C) failed to produce diffraction-quality crystals of wild-type  $\beta_2$ AR.

## Challenges in crystallizing GPCRs

The difficulty in generating crystals from the wild-type  $\beta_2$ AR and other GPCRs for diffusible hormones and neurotransmitters may be related to the observation that these molecules are conformationally complex<sup>18,21,22</sup>. In contrast to rhodopsin, many GPCRs, including the  $\beta_2$ AR, exhibit significant basal, agonist-independent G protein activation. This basal activity has been associated with structural instability<sup>23,24</sup>, suggesting that the intramolecular interactions that maintain the receptor in the inactive state are also important for the structural integrity of the protein. Orthosteric ligands for GPCRs exhibit a spectrum of efficacies for receptor-stimulated G protein activation, ranging from inverse agonists, which inhibit basal activity, to agonists, which maximally activate the receptor.

The  $\beta_2$ AR contains relatively unstructured regions that are involved in functionally important protein–protein interactions. Protease susceptibility and intramolecular fluorescence resonance energy transfer experiments<sup>25</sup> indicate that the C terminus and the third intracellular loop are the most unstructured regions. The N and C-terminal ends of the third intracellular loop are involved in G protein activation and the selectivity of GPCR–G protein interactions<sup>26</sup>. The C terminus interacts with G-protein-coupled receptor kinases, arrestins and other signalling molecules<sup>27</sup>. In the case of water-soluble proteins, removal of such unstructured regions can facilitate crystallization, but for the  $\beta_2$ AR this strategy would remove hydrophilic surfaces that are frequently observed to form lattice contacts in membrane-protein crystals.

## Crystallization and structure solution

In an effort to provide conformational stability while increasing the polar surface available for crystal contacts, we generated a monoclonal antibody (Mab5) that binds to the third intracellular loop of native, but not denatured receptor protein<sup>28</sup>. Mab5 was generated

<sup>1</sup>Department of Molecular and Cellular Physiology and <sup>2</sup>Department of Structural Biology, Stanford University School of Medicine, 279 Campus Drive, Stanford, Palo Alto, California 94305, USA. <sup>3</sup>MRC Laboratory of Molecular Biology, Cambridge CB2 2QH, UK. <sup>4</sup>European Synchrotron Radiation Facility, 6 rue Jules Horowitz, BP220, 38043 Grenoble, cedex 9, France. <sup>5</sup>Biosciences Division, Argonne National Laboratory, GM/CA-CAT, Boulevard 436, D007, 9700 South Cass Avenue, Argonne, Illinois 60439, USA.

\*These authors contributed equally to this work.

by immunizing mice with purified  $\beta_2$ AR reconstituted into phospholipid vesicles at a high protein-to-lipid ratio. Binding of Mab5 to  $\beta_2$ AR does not alter agonist or antagonist binding affinities, and does not prevent agonist-induced conformational changes<sup>28</sup>; therefore, it does not significantly alter the native structure of the receptor. Purified, deglycosylated  $\beta_2$ AR bound to carazolol (an inverse agonist) forms a complex with the Fab generated from Mab5 (Fab5) in detergent, and the  $\beta_2$ AR–Fab5 complex can be isolated by size-exclusion chromatography.

Crystals of the carazolol-bound  $\beta_2$ AR–Fab5 complex were grown in DMPC bicelles<sup>29</sup> using ammonium sulphate as a precipitant. The size and uniformity of the crystals were improved by removing 48 amino acids from the unstructured C terminus ( $\beta_2$ AR365, Fig. 1). Crystals of the  $\beta_2$ AR365–Fab5 complex grew as long, thin plates up to 300- $\mu$ m long, approximately 30- $\mu$ m wide, and less than 10- $\mu$ m thick. Owing to the size and radiation sensitivity of the crystals, data collection required the use of microbeam technology<sup>7,30</sup> in which X-ray beams are either focused (ID-13 and ID23-2 beamlines, European Synchrotron Radiation Facility, Grenoble) or moderately focused and then further collimated (23ID-B GM/CA-CAT beamline, Advanced Photon Source) to diameters between 5 and 10  $\mu$ m. The initial images from the best crystals showed diffraction to 3.0 Å; however, resolution was rapidly lost in sequential images from the same crystal volume. Nevertheless, we obtained a complete data set from a single crystal, and determined the structure by molecular replacement using immunoglobulin-domain search models for the Fab. The diffraction is anisotropic, with diffraction extending to 3.4 Å in the plane of the membrane and 3.7 Å perpendicular to the plane of the membrane.

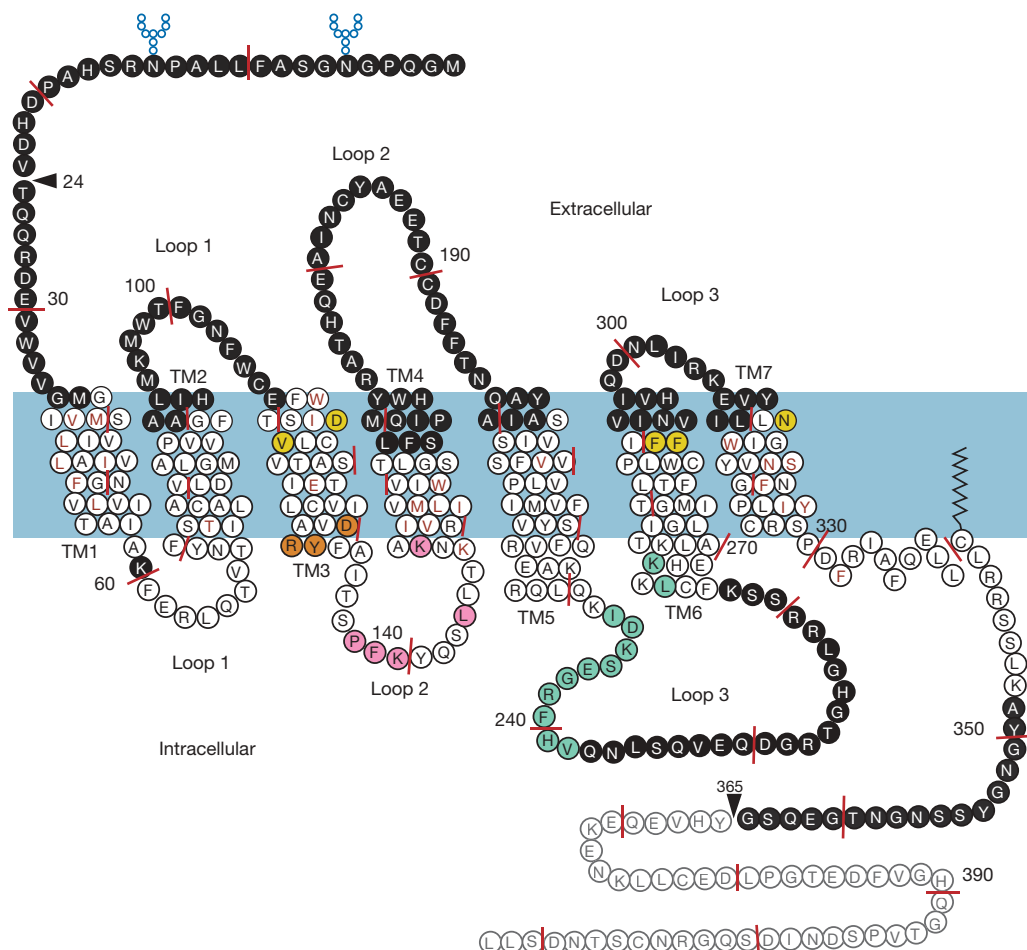
### Structure of the $\beta_2$ AR–Fab5 complex

Figure 2a shows the packing of the  $\beta_2$ AR365–Fab5 complex in the crystals. The crystals seem to be formed from stacks of two-dimensional

crystals, as previously reported for bacteriorhodopsin crystallized in bicelles<sup>31</sup>. There are few contacts between adjacent receptor molecules within a bicelle layer, indicating that the receptor is monomeric in the crystal. This is somewhat surprising considering that, in all reported crystals of rhodopsin, rhodopsin exists as antiparallel or parallel dimers<sup>1–7</sup>. Moreover, evidence from a variety of biochemical and biophysical studies suggest that the  $\beta_2$ AR and many other GPCRs exist as dimers or higher-order oligomers in the plasma membrane of cultured cells<sup>32</sup>, and there may be a role for dimers in the export of properly folded receptor protein from the endoplasmic reticulum<sup>32</sup>. It is important to note, however, that  $\beta_2$ AR dimerization is not required for G protein activation. Purified  $\beta_2$ AR exists as monomers, and monomeric  $\beta_2$ AR reconstituted into recombinant high-density lipoprotein particles couples efficiently to  $G_s$ —its preferred heterotrimeric G protein<sup>33</sup>.

The best-resolved regions of the crystal are the Fab5 fragments and cytoplasmic ends of the transmembrane segments of the receptor (Supplementary Fig. 1). In contrast to the cytoplasmic side of the receptor, the electron density is uninterpretable in the extracellular domain (Supplementary Fig. 2), even though this region of two receptor molecules packs together in a head-to-head manner around the crystallographic two-fold axis. The poor packing in this interface probably explains the significant anisotropy and poor overall resolution of the crystals. In an effort to improve the packing of the extracellular domains, we further modified  $\beta_2$ AR365 by inserting a TEV cleavage site after amino acid 24 ( $\beta_2$ AR24/365, Fig. 1). However, crystals of this construct are isomorphous to those made with  $\beta_2$ AR365, and the structure (Supplementary Table 1) is virtually identical to that obtained from  $\beta_2$ AR365–Fab5.

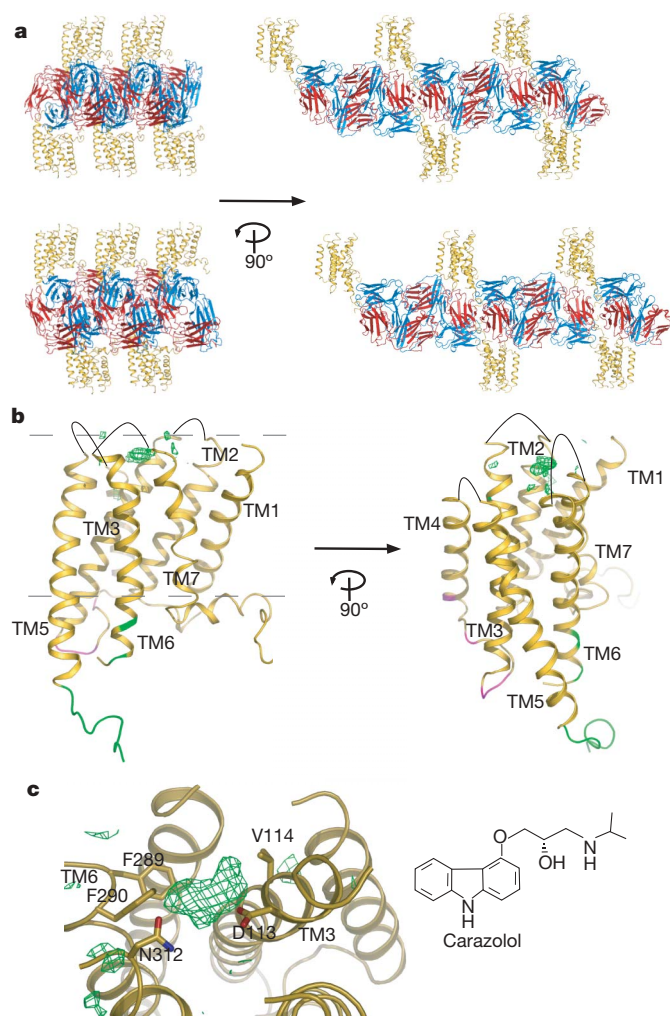
As expected, the overall structure of the  $\beta_2$ AR (Fig. 2b) is similar to rhodopsin, with seven transmembrane helices and an eighth helix that runs parallel to the cytoplasmic face of the membrane. Several of



**Figure 1 | Schematic diagram of the  $\beta_2$ AR.** Black circles with white letters indicate disordered residues not included in the model. Grey letters and circles indicate residues not included in the  $\beta_2$ AR365 construct used for crystallography. Red letters indicate amino acids for which side-chain electron density was not modelled. Yellow residues indicate amino acids implicated in ligand binding from mutagenesis studies. Orange residues indicate the conserved DRY sequence. Green residues form the Fab5 epitope, and pink residues are packed against the Fab5 constant domain in the lattice. Small blue circles indicate glycosylation sites. Red lines indicate ten-amino-acid increments.



the transmembrane helices are broken by non-helical kinks, most prominently TM7. Residues not included in the  $\beta_2$ AR model, owing to absent or uninterpretable electron density, are indicated in Fig. 1. In the transmembrane helices, the majority of the missing side chains face the lipid environment. The loss of electron density occurs just above the ligand-binding site, near the predicted lipid-water interface, suggesting that ligand binding and/or the lipid environment contributes to the order of the transmembrane segments. Specific interactions between the variable domains of Fab5 and the  $\beta_2$ AR occur over a sequence of nine amino acids at the N-terminal end of intercellular loop 3 (I233–V242) and two amino acids at the C-terminal end (L266 and K270) (shown in green in Fig. 2b). Therefore, Fab5 recognizes a three-dimensional epitope on the  $\beta_2$ AR, which is in agreement with the observation that Fab5 binds to native, but not denatured  $\beta_2$ AR protein<sup>28</sup>. Additional lattice contacts occur between the constant domain of a symmetry-related Fab5 molecule and the second intracellular loop of  $\beta_2$ AR (shown in magenta in Fig. 2b).



**Figure 2 | Structure of the  $\beta_2$ AR365–Fab5 complex.** **a**, Packing of the  $\beta_2$ AR365–Fab5 complex in crystals formed in DMPC bicelles ( $\beta_2$ AR, gold; heavy chain, blue; light chain, red). **b**, Structure of the  $\beta_2$ AR showing sites of the interactions with Fab5. Sites of specific (idiotypic) interactions between Fab5 and the  $\beta_2$ AR are shown in green. Sites of interactions between the  $\beta_2$ AR and the constant region of Fab5 of the symmetry mate are shown in magenta. Dotted grey lines indicate predicted membrane boundaries. Solid black lines indicate extracellular connections between transmembrane segments. **c**,  $F_O - F_C$  map contoured at  $2.0 \sigma$  and surrounded by residues known to be involved in ligand binding. The chemical structure of carazolol, the bound ligand, is shown on the right.

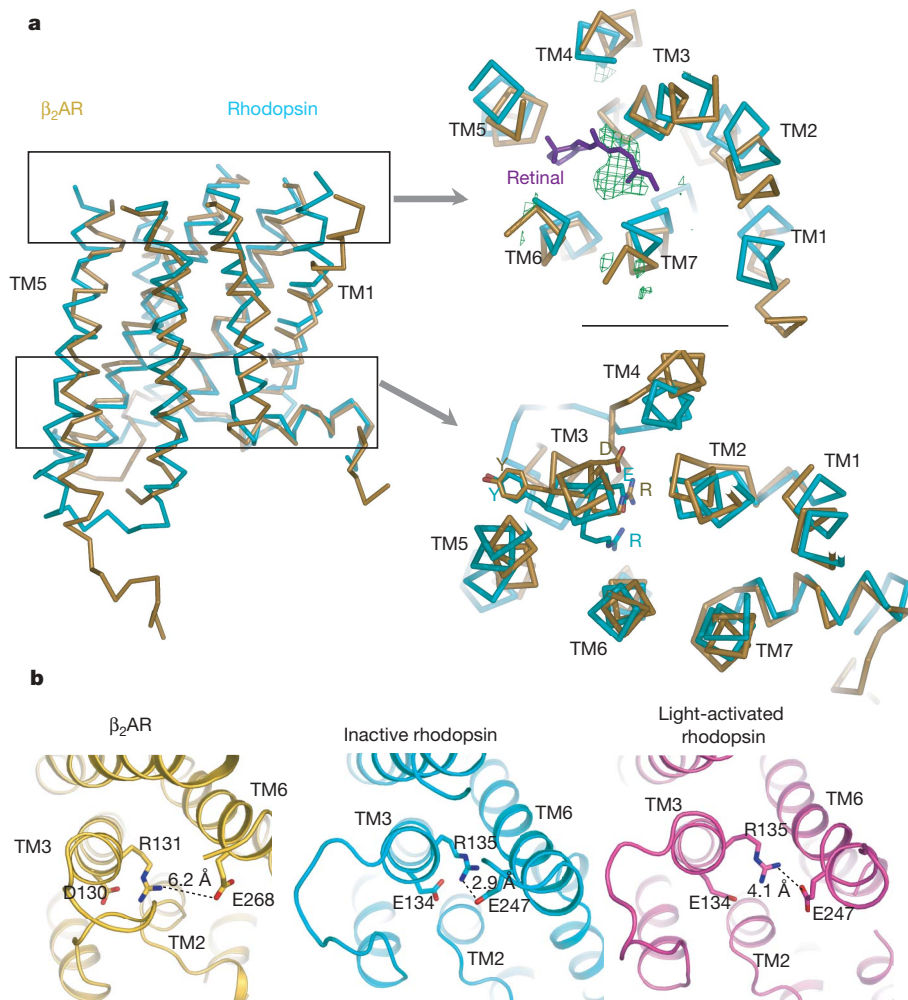
## Structural insights into basal activity

The ligand-binding site can be identified by an extended flat feature in the electron-density maps close to the extracellular side of the transmembrane helices (Fig. 2b, c). This is the only large feature in residual electron-density maps and is adjacent to Asp 113, Val 114, Phe 289, Phe 290 and Asn 312—residues identified from mutagenesis studies as being involved in ligand binding in the  $\beta_2$ AR<sup>9,34,35</sup>. This region corresponds to the retinal-binding site of rhodopsin. The weak electron density in this region precludes definitive modelling of carazolol. It is unlikely that the crystallization conditions resulted in dissociation of carazolol from  $\beta_2$ AR. Carazolol bound to the  $\beta_2$ AR has a distinct fluorescence emission spectrum<sup>36</sup>, and  $\beta_2$ AR crystals and associated protein precipitate harvested from equilibrated hanging-drops showed no significant loss of carazolol binding, as detected by fluorescence spectroscopy (data not shown).

Figure 3 shows a comparison of transmembrane segments of the  $\beta_2$ AR superimposed with the homologous structure of rhodopsin. The root mean squared deviation for the alpha carbon backbone of the transmembrane segments is 1.56 Å. Although the overall arrangement of the transmembrane segments is similar, the  $\beta_2$ AR has a more open structure. The difference in the arrangement of the cytoplasmic ends of the transmembrane segments of  $\beta_2$ AR and rhodopsin may provide structural insights into basal receptor activity. Rhodopsin has no detectable basal activity, a feature essential for vision. In contrast, even when bound to the inverse agonist carazolol, the comparatively high basal activity of the  $\beta_2$ AR is suppressed by only 50% (Supplementary Fig. 4). Therefore, the carazolol bound  $\beta_2$ AR is not functionally equivalent to dark rhodopsin. Figure 3b compares the  $\beta_2$ AR and two rhodopsin structures at the level of the conserved (E/D)R(Y/W) sequence (found in 72% of rhodopsin family members)<sup>18</sup>. In the high-resolution structure of inactive (dark) rhodopsin, E134 and R135 in TM3 and E247 in TM6 form a network of hydrogen bonds and charge interactions referred to as the ‘ionic lock’<sup>37</sup>. These interactions maintain rhodopsin in an inactive conformation. The ionic lock residues seem to have a similar role in the  $\beta_2$ AR because mutations of these amino acids in the  $\beta_2$ AR or other adrenergic receptors lead to constitutive activity<sup>37,38</sup>. Moreover, evidence from biophysical studies suggests that movement of the cytoplasmic end of TM3 relative to TM6 on activation is similar for the  $\beta_2$ AR and rhodopsin<sup>17,39</sup>. However, as shown in Fig. 3b, the transmembrane segments of the  $\beta_2$ AR have a more open structure in this region, and R131 in carazolol-bound  $\beta_2$ AR is not close enough to E268 to form a hydrogen bond. The structure of carazolol-bound  $\beta_2$ AR around the ionic lock is more similar to the structure of light-activated rhodopsin<sup>40</sup> (Fig. 3b), in which R135 and E247 are separated by 4.1 Å. This light-activated rhodopsin structure may not represent the fully active conformation because the spectral properties of these crystals are similar, but not identical, to those of metarhodopsin II<sup>40</sup>. Nevertheless, given the role of TM3, TM6 and the adjacent cytoplasmic loops in G protein coupling, the more open structure of the  $\beta_2$ AR may account for the residual basal activity of the  $\beta_2$ AR bound to the inverse agonist carazolol.

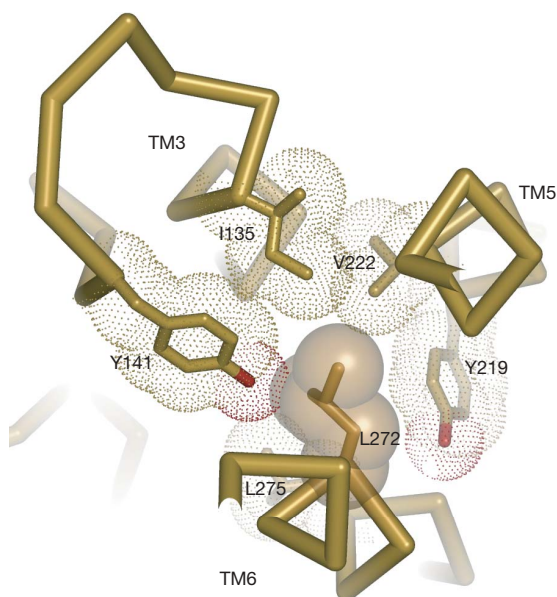
It is unlikely that the observed structural differences between the  $\beta_2$ AR and rhodopsin are due to distortion of the  $\beta_2$ AR owing to interactions between Fab5 and the third intracellular loop, because binding of Fab5 had no effect on agonist or antagonist binding affinity, and does not effect agonist-induced movement of TM3 relative to TM6 (ref. 28). However, we cannot exclude the possibility that crystal packing interactions between Fab5 and the second extracellular loop (Fig. 2b) contribute to these structural differences.

Another set of intramolecular interactions known to be important for minimizing the basal activity of the  $\beta_2$ AR involves L272 in TM6. Mutation of L272 to alanine was the first reported constitutively active mutant of the  $\beta_2$ AR<sup>41</sup>. As seen in Fig. 4, L272 forms extensive van der Waals interactions with I135 in TM3; V222 and Y219 in TM5; and Y141 in intracellular loop 2 (Fig. 4, and Supplementary Fig. 5). Because L272 is adjacent to E268, disruption of the packing interactions by mutation to alanine may have an effect similar to



**Figure 3 | Comparison of  $\beta_2$ AR and rhodopsin structures.** **a**, The  $\beta_2$ AR is superimposed with the homologous structure of rhodopsin<sup>6</sup>. Retinal is shown in purple and the electron density in the putative ligand-binding site is shown as a green mesh. Structures were aligned using all seven transmembrane segments. The right panels represent cross-sections that are rotated 90° around the horizontal axis and viewed from the extracellular face of the receptor. **b**, Comparison of the  $\beta_2$ AR with structures of inactive rhodopsin and light-activated rhodopsin around the conserved E/DRY sequence in TM3. A dashed line shows the distance between the homologous arginine in TM3 and glutamate in TM6. To facilitate comparison of the E/DRY regions, the structures were aligned by superimposing TM3 only.

disruption of the ionic lock in rhodopsin. It is likely that this mutation would produce a more loosely packed, dynamic structure in this region, shifting the equilibrium towards a more active state.



**Figure 4 | Side-chain interactions between Leu 272 and residues in TM3, TM5 and intracellular loop 2.** Packing interactions are reflected in lower *B*-factors for these amino acids. The average *B* value of residues 135, 141, 219, 222, 272 and 275 is 117 Å<sup>2</sup>, compared to 157 Å<sup>2</sup> for the receptor as a whole.

It is interesting that packing interactions around L272 are observed while the ionic lock interactions are absent. Because mutation of either E268 or L272 leads to elevated basal activity, it is likely that both are involved in maintaining the basal state of the receptor. From the current structure, we can conclude that formation of the ionic lock and the tight packing of L272 are not interdependent, and might even be structurally incompatible. It is possible that the ionic lock and L272 interactions stabilize two of several distinct substates in the unliganded  $\beta_2$ AR, and that these two substates have lower activity towards G<sub>s</sub> than the others. Carazolol binding may further stabilize the substate that favours packing around L272, and therefore reduce basal activity relative to the ensemble of substates in the unliganded receptor. The residual activity in the carazolol-bound receptor may be due to the failure to stabilize ionic lock interactions.

The limitations of this crystal structure of the  $\beta_2$ AR can be attributed to the poor crystal packing and the inherent structural flexibility of this GPCR relative to rhodopsin. Different crystallographic approaches will be needed to stabilize and visualize the extracellular domain and provide a more detailed picture of extracellular loops as well as the ligand-binding site. Nevertheless, this structure of the  $\beta_2$ AR in a lipid environment provides structural insights into the basis of basal activity, a feature of many GPCRs that may have both physiologic and therapeutic relevance.

#### METHODS SUMMARY

$\beta_2$ AR was expressed in Sf9 insect cells using recombinant baculovirus. Sf9 cell membranes were solubilized in dodecylmaltoside and purified by sequential antibody and ligand affinity chromatography. Fab5 was generated by papain digestion of Mab5 and purified by ion-exchange chromatography. The  $\beta_2$ AR–Fab5 complex was formed by mixing purified  $\beta_2$ AR with a stoichiometric



excess of Fab5, and then isolated by size-exclusion chromatography. The purified  $\beta_2$ AR–Fab5 complex was mixed with bicelles composed of the lipid DMPC and the detergent CHAPSO. The final  $\beta_2$ AR–Fab concentration ranged between 8 and 12 mg ml<sup>-1</sup>. Crystals were grown by hanging-drop vapour diffusion in a mixture of ammonium sulphate, sodium acetate and EDTA over a pH range of 6.5 to 7.5. Crystals grew within 7 to 10 days. They were cryoprotected in 20% glycerol before freezing in liquid nitrogen. Owing to the size and radiation sensitivity of the crystals, diffraction images were obtained by microcrystallography. The structure of the  $\beta_2$ AR365–Fab5 complex was solved by molecular replacement, using separate constant and variable Fab domain structures as search models. Coordinates and structure factors are deposited in the Protein Data Bank (accession codes 2R4R for  $\beta_2$ AR365–Fab5 and 2R4S for  $\beta_2$ AR24/365–Fab5).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 26 July; accepted 28 September 2007.**

**Published online 21 October 2007.**

- Okada, T. *et al.* X-Ray diffraction analysis of three-dimensional crystals of bovine rhodopsin obtained from mixed micelles. *J. Struct. Biol.* **130**, 73–80 (2000).
- Palczewski, K. *et al.* Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**, 739–745 (2000).
- Okada, T. *et al.* Functional role of internal water molecules in rhodopsin revealed by X-ray crystallography. *Proc. Natl Acad. Sci. USA* **99**, 5982–5987 (2002).
- Teller, D. C., Okada, T., Behnke, C. A., Palczewski, K. & Stenkamp, R. E. Advances in determination of a high-resolution three-dimensional structure of rhodopsin, a model of G-protein-coupled receptors (GPCRs). *Biochemistry* **40**, 7761–7772 (2001).
- Okada, T. *et al.* The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. *J. Mol. Biol.* **342**, 571–583 (2004).
- Li, J., Edwards, P. C., Burghammer, M., Villa, C. & Schertler, G. F. Structure of bovine rhodopsin in a trigonal crystal form. *J. Mol. Biol.* **343**, 1409–1438 (2004).
- Standfuss, J. *et al.* Crystal structure of a thermally stable rhodopsin mutant. *J. Mol. Biol.* **372**, 1179–1188 (2007).
- Lefkowitz, R. J. The superfamily of heptahelical receptors. *Nature Cell Biol.* **2**, E133–E136 (2000).
- Strader, C. D., Sigal, I. S. & Dixon, R. A. Structural basis of  $\beta$ -adrenergic receptor function. *FASEB J.* **3**, 1825–1832 (1989).
- Wieland, K., Zuurmond, H. M., Krasel, C., Ijzerman, A. P. & Lohse, M. J. Involvement of Asn-293 in stereospecific agonist recognition and in activation of the  $\beta_2$ -adrenergic receptor. *Proc. Natl Acad. Sci. USA* **93**, 9276–9281 (1996).
- Liapakis, G. *et al.* The forgotten serine. A critical role for Ser-203<sup>5,42</sup> in ligand binding to and activation of the  $\beta_2$ -adrenergic receptor. *J. Biol. Chem.* **275**, 37779–37788 (2000).
- Strader, C. D., Candelore, M. R., Hill, W. S., Sigal, I. S. & Dixon, R. A. F. Identification of two serine residues involved in agonist activation of the  $\beta$  adrenergic receptor. *J. Biol. Chem.* **264**, 13572–13578 (1989).
- Ghanouni, P., Steenhuis, J. J., Farrens, D. L. & Kobilka, B. K. Agonist-induced conformational changes in the G-protein-coupling domain of the  $\beta_2$  adrenergic receptor. *Proc. Natl Acad. Sci. USA* **98**, 5997–6002 (2001).
- Swaminath, G. *et al.* Sequential binding of agonists to the  $\beta_2$  adrenoceptor: kinetic evidence for intermediate conformational states. *J. Biol. Chem.* **279**, 686–691 (2004).
- Ghanouni, P. *et al.* Functionally different agonists induce distinct conformations in the G protein coupling domain of the  $\beta_2$  adrenergic receptor. *J. Biol. Chem.* **276**, 24433–24436 (2001).
- Swaminath, G. *et al.* Probing the  $\beta_2$  adrenoceptor binding site with catechol reveals differences in binding and activation by agonists and partial agonists. *J. Biol. Chem.* **280**, 22165–22171 (2005).
- Yao, X. *et al.* Coupling ligand structure to specific conformational switches in the  $\beta_2$ -adrenoceptor. *Nature Chem. Biol.* **2**, 417–422 (2006).
- Kobilka, B. K. & Deupi, X. Conformational complexity of G-protein-coupled receptors. *Trends Pharmacol. Sci.* **28**, 397–406 (2007).
- Kobilka, B. K. Amino and carboxyl terminal modifications to facilitate the production and purification of a G protein-coupled receptor. *Anal. Biochem.* **231**, 269–271 (1995).
- Caron, M. G., Srinivasan, Y., Pitha, J., Kocielek, K. & Lefkowitz, R. J. Affinity chromatography of the  $\beta$ -adrenergic receptor. *J. Biol. Chem.* **254**, 2923–2927 (1979).
- Kenakin, T. Ligand-selective receptor conformations revisited: the promise and the problem. *Trends Pharmacol. Sci.* **24**, 346–354 (2003).
- Kobilka, B. K. G protein coupled receptor structure and activation. *Biochim. Biophys. Acta* **1768**, 794–807 (2006).
- Gether, U. *et al.* Structural instability of a constitutively active G protein-coupled receptor. Agonist-independent activation due to conformational flexibility. *J. Biol. Chem.* **272**, 2587–2590 (1997).
- Samama, P., Bond, R. A., Rockman, H. A., Milano, C. A. & Lefkowitz, R. J. Ligand-induced overexpression of a constitutively active  $\beta_2$ -adrenergic receptor: Pharmacological creation of a phenotype in transgenic mice. *Proc. Natl Acad. Sci. USA* **94**, 137–141 (1997).
- Granier, S. *et al.* Structure and conformational changes in the C-terminal domain of the  $\beta_2$ -adrenoceptor: insights from fluorescence resonance energy transfer studies. *J. Biol. Chem.* **282**, 13895–13905 (2007).
- Gether, U. & Kobilka, B. K. G protein-coupled receptors. II. Mechanism of agonist activation. *J. Biol. Chem.* **273**, 17979–17982 (1998).
- Reiter, E. & Lefkowitz, R. J. GRKs and  $\beta$ -arrestins: roles in receptor silencing, trafficking and signaling. *Trends Endocrinol. Metab.* **17**, 159–165 (2006).
- Day, P. W. *et al.* A monoclonal antibody for G protein coupled receptor crystallography. *Nature Methods* doi:10.1038/nmeth1112 (21 October 2007).
- Faham, S. *et al.* Crystallization of bacteriorhodopsin from bicelle formulations at room temperature. *Protein Sci.* **14**, 836–840 (2005).
- Riekel, C., Burghammer, M. & Schertler, G. Protein crystallography microdiffraction. *Curr. Opin. Struct. Biol.* **15**, 556–562 (2005).
- Faham, S. & Bowie, J. U. Bicelle crystallization: a new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure. *J. Mol. Biol.* **316**, 1–6 (2002).
- Bulenger, S., Marullo, S. & Bouvier, M. Emerging role of homo- and heterodimerization in G-protein-coupled receptor biosynthesis and maturation. *Trends Pharmacol. Sci.* **26**, 131–137 (2005).
- Whorton, M. R. *et al.* A monomeric G protein-coupled receptor isolated in a high-density lipoprotein particle efficiently activates its G protein. *Proc. Natl Acad. Sci. USA* **104**, 7682–7687 (2007).
- Suryanarayana, S., Daunt, D. A., Von Zastrow, M. & Kobilka, B. K. A point mutation in the seventh hydrophobic domain of the  $\alpha_2$  adrenergic receptor increases its affinity for a family of  $\beta$  receptor antagonists. *J. Biol. Chem.* **266**, 15488–15492 (1991).
- Chelikani, P., Hornak, V., Eilers, M., Reeves, P. J., Smith, S. O., RajBhandary, U. L. & Khorana, H. G. Role of group-conserved residues in the helical core of  $\beta_2$ -adrenergic receptor. *Proc. Natl Acad. Sci. USA* **104**, 7027–7032 (2007).
- Tota, M. R. & Strader, C. D. Characterization of the binding domain of the  $\beta$ -adrenergic receptor with the fluorescent antagonist carazolol. Evidence for a buried ligand binding site. *J. Biol. Chem.* **265**, 16891–16897 (1990).
- Ballesteros, J. A. *et al.* Activation of the  $\beta_2$ -adrenergic receptor involves disruption of an ionic lock between the cytoplasmic ends of transmembrane segments 3 and 6. *J. Biol. Chem.* **276**, 29171–29177 (2001).
- Scheer, A. *et al.* Mutational analysis of the highly conserved arginine within the Glu/Asp-Arg-Tyr motif of the  $\alpha_{1B}$ -adrenergic receptor: effects on receptor isomerization and activation. *Mol. Pharmacol.* **57**, 219–231 (2000).
- Farrens, D. L., Altenbach, C., Yang, K., Hubbell, W. L. & Khorana, H. G. Requirement of rigid-body motion of transmembrane helices for light activation of rhodopsin. *Science* **274**, 768–770 (1996).
- Salom, D. *et al.* Crystal structure of a photoactivated deprotonated intermediate of rhodopsin. *Proc. Natl Acad. Sci. USA* **103**, 16123–16128 (2006).
- Samama, P., Cotecchia, S., Costa, T. & Lefkowitz, R. J. A mutation-induced activated state of the  $\beta_2$ -adrenergic receptor. Extending the ternary complex model. *J. Biol. Chem.* **268**, 4625–4636 (1993).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This study was supported by the Lundbeck Foundation (S.G.F.R.), a National Institutes of Health Ruth L. Kirchstein NRSA grant (D.M.R.), a National Institute of General Medical Sciences grant (W.I.W.), a National Institute of Neurological Disorders and Stroke grant, the Mather Charitable Foundation, and a generous gift from Lundbeck (to B.K.K.). G.F.X.S. was financially supported by a Human Frontier Science Project (HFSP) programme grant, a European Commission FP6 specific targeted research project and an ESRF long-term proposal. We thank R. Mackinnon and J. Bowie for advice, R. Stevens for help with early screening efforts, and J. Smith for arranging access to GM/CA-CAT at the APS. Use of the APS is supported by the US Department of Energy. GM/CA-CAT is funded by the US National Institutes of Cancer and General Medical Sciences. We thank X. Deupi and S. Granier for help with data collection. We thank D. Flot for his support at the ID 23.2 microfocus beamline at the European Synchrotron Radiation Facility.

**Author Contributions** S.G.F.R. performed final stages of  $\beta_2$ AR purification, purified Mab5 and prepared Fab5. D.M.R. generated recombinant  $\beta_2$ AR used for crystallography. Crystal screening and optimization were performed by S.G.F.R. and D.M.R. H.J.C. assisted with data collection at the Advanced Photon Source, processed all diffraction data and solved the structure of the  $\beta_2$ AR–Fab5 complex. F.S.T. expressed  $\beta_2$ AR in insect cells and, together with T.S.K., performed the initial stage of  $\beta_2$ AR purification. T.S.K. prepared antibody 5. W.I.W. supervised and assisted with data collection at the Advanced Photon Source, and with data processing and structure determination. G.F.X.S. introduced B.K.K. to microfocus diffraction technology and supervised data collection at the European Synchrotron Radiation Facility. P.C.E. and M.B. assisted with data collection at the European Synchrotron Radiation Facility. R.S. and R.F.F. assisted with data collection at the Advanced Photon Source. V.R.P.R. performed the functional characterization of carazolol. B.K.K. was responsible for the overall project management and strategy, and assisted with  $\beta_2$ AR purification, crystal harvesting and synchrotron data collection. B.K.K., W.I.W. and G.F.X.S. prepared the manuscript. All authors discussed the results and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to B.K.K. ([kobilka@stanford.edu](mailto:kobilka@stanford.edu)).



## METHODS

**Crystallization.** Preparation of  $\beta_2$ AR365 and Fab5 are described in Supplementary Methods. The  $\beta_2$ AR365–Fab5 complexes were mixed with bicelles (10% w/v 3:1 DMPC:CHAPSO in 10 mM HEPES, pH 7.5, 100 mM NaCl) at a 1:5 (protein:bicelle) ratio, and crystals were grown in sitting- and hanging-drop formats at 22 °C using equal volumes of protein mixture and reservoir solutions. Initial crystallization leads were identified using multiple 96-well sitting-drop screens from Nextal (Qiagen). After extensive optimization, crystals for data collection were grown in hanging-drop format over a reservoir solution of 1.85–2.0 M ammonium sulphate, 180 mM sodium acetate, 5 mM EDTA, 100 mM MES or HEPES, pH 6.5–7.5. Crystals grew to full size within 7 to 10 days. Crystals were flash frozen and stored in liquid nitrogen, with reservoir solution plus 20% glycerol as cryoprotectant.

**Microcrystallography data collection and processing.** Microbeams were essential to obtain a favourable signal-to-noise ratio from the weakly diffracting thin crystals. The shape of the crystals permitted complete data to be measured from a single crystal. A small wedge of data, typically 5–10°, (1° per frame) could be measured before significant radiation damage was observed. The crystal was then translated to a new, undamaged position to collect the next wedge of data. A total of 182° of data collected in this manner, measured at beamline ID23-2 of the ESRF, were used for the final  $\beta_2$ AR365–Fab5 data set (Supplementary Table 1). The  $\beta_2$ AR24/365–Fab5 data set was obtained from 225° of data measured using a 4- $\mu$ m  $\times$  6- $\mu$ m beam at beamline 23ID-B of the APS (Supplementary Table 1).

ESRF data were processed with MOSFLM and SCALA<sup>42</sup>, and data measured at the APS were processed with HKL2000<sup>43</sup>. In many cases it was necessary to re-index the crystal after moving to a new position on the crystal, which may have been due to bending of the frozen crystals such that the indexing matrix from the previous volume could not accurately predict the diffraction pattern from a new volume. This problem precluded global post refinement of the unit cell parameters. The unit cell parameters used for subsequent analysis (Supplementary Table 1) were obtained from initial indexing and refinement from one wedge of the ESRF data, and were subsequently found to be sufficient for processing the remaining data without unit cell constant refinement. Using a partial specific volume of 1.21 Å<sup>3</sup>/Da for protein, the unit cell would have 66% lipid, detergent and aqueous solvent for one  $\beta_2$ AR–Fab5 complex in the asymmetric unit.

**Structure solution and refinement.** The structure of the  $\beta_2$ AR365–Fab5 complex was solved by molecular replacement, by searching with separate constant and variable domain models against a low-resolution (4.1 Å) data set measured at ESRF beamline ID-13. The Fab was derived from a murine IgG antibody containing a  $\kappa$  light chain and  $\gamma$ 1 heavy chain<sup>28</sup>. At the time of these calculations the sequence of the heavy chain was not known, and the crystal structure of a Fab containing a  $\kappa$  light chain but  $\gamma$ 2 heavy chain<sup>44</sup> (PDB code 1IGT) was used as a search model. Molecular replacement was performed with the program PHASER<sup>45</sup>, using data between 12 and 4.5 Å. The constant domain was placed first, followed by the variable domain. The constant domain model retained all side chains, whereas the variable domain was reduced to polyalanine. All atomic temperature factors were set to 50 Å<sup>2</sup>. The best solution had rotation and translation function *Z* scores of 5.3 and 10.6 for the constant domain, and 4.5 and 21.7 for the variable domain. An electron density map calculated to 6 Å from this solution revealed rods of density corresponding to the transmembrane helices of the receptor. A model of the transmembrane portion of rhodopsin made by removing the cytoplasmic and extracellular loops, retinal and water molecules, and replacing those residues non-identical with  $\beta_2$ AR with alanine could be manually placed into this density. To obtain a convenient starting model for

building the receptor, the molecular replacement calculation was re-run to include the rhodopsin transmembrane helices model as a third search model after placing the two Fab domains. Although the top solution was not very strong statistically (rotation function *Z* = 2.5, translation function *Z* = 7.0), after rigid body refinement the rhodopsin model was very close to that placed manually into the 6 Å map. This molecular replacement solution was then subjected to rigid-body refinement between 20 and 5 Å in CNS<sup>46</sup>, using five rigid bodies (the Fab constant domain light and heavy chains, the variable domain light and heavy chains, and rhodopsin). This gave *R* and *R*<sub>free</sub> values of 0.447 and 0.452, respectively.

Electron-density maps made with phases either from the Fab model alone or the rigid-body refined Fab + minimal rhodopsin model indicated significant differences between rhodopsin and  $\beta_2$ AR, and extensive manual rebuilding was required to refine the structure. The structure was initially refined at 4.1 Å resolution. The test set from the 4.1 Å set was transferred to the higher-resolution  $\beta_2$ AR365–Fab5 set measured at the ESRF (Supplementary Table 1) and additional test set reflections added in the 4.1–3.4 Å range. Multiple rounds of manual rebuilding, positional and grouped temperature factor refinement were performed using the maximum likelihood amplitude target in CNS. The electron density of the Fab is very well defined owing to its tight packing in the crystal, whereas the receptor is poorly packed and has much higher temperature factors (Supplementary Fig. 1 and Supplementary Table 1). Because the receptor density is poor, we also refined against a second data set from a single crystal of the  $\beta_2$ AR24/365–Fab5 complex (Supplementary Table 1), to ensure that any densities observed in the receptor region are not due to noise in the first data set. The  $\beta_2$ AR24/365–Fab5 data set was obtained from 225° of data measured using a 4- $\mu$ m  $\times$  6- $\mu$ m beam at beamline 23ID-B of the APS. Although there is electron density in the extracellular region, the final model retains only those residues that could be unambiguously assigned (Fig. 1).

The high-temperature factors and weak electron density for the receptor raises concerns about model bias. However, the Fab represents 50% of the scattering mass and, because of its better order, contributes even more to the total scattering and so represents a significant source of phase information independent of the receptor. Simulated annealing omit maps confirmed the interpretation presented here. Moreover, alternative sequence registers or backbone paths were considered in several portions of the receptor, but these models could be eliminated based on inspection of  $\sigma_A$  weighted  $2F_O - F_C$  and  $F_O - F_C$  electron density maps.

On the basis of the average  $F/\sigma(F)$  of reflections near the three crystallographic axes (as defined by the program TRUNCATE<sup>32</sup>), we estimate the effective resolution to be 3.4 Å within the plane of the membrane and 3.7 Å perpendicular to the membrane for the  $\beta_2$ AR365–Fab5 structure, and 3.4 Å/3.8 Å for the  $\beta_2$ AR24/365–Fab5 structure.

42. Collaborative Computational Project. N. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
43. Otwinowski, Z. & Minor, W. Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
44. Harris, L. J., Larson, S. B., Hasel, K. W. & McPherson, A. Refined structure of an intact IgG2a monoclonal antibody. *Biochemistry* **36**, 1581–1597 (1997).
45. McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D* **63**, 32–41 (2007).
46. Brünger, A. T. *et al.* Crystallography and NMR System (CNS): A new software system for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).

## LETTERS

# A runaway collision in a young star cluster as the origin of the brightest supernova

Simon F. Portegies Zwart<sup>1,2</sup> & Edward P. J. van den Heuvel<sup>1,3</sup>

Supernova SN 2006gy in the galaxy NGC 1260 is the most luminous recorded<sup>1–4</sup>. Its progenitor might have been a very massive ( $>100M_{\odot}$ , where  $M_{\odot}$  is the mass of the Sun) star<sup>5</sup>, but that interpretation is incompatible with hydrogen in the spectrum of the supernova; stars  $>40M_{\odot}$  are believed to have shed their hydrogen envelopes several hundred thousand years before the explosion<sup>6</sup>. Alternatively, the progenitor might have arisen from the merger of two massive stars<sup>7</sup>. Here we show that the collision frequency of massive stars in a dense and young cluster (of the kind to be expected near the centre of a galaxy) is sufficient to provide a reasonable chance that SN 2006gy resulted from such a bombardment. If this is the correct explanation, then we predict that when the supernova fades (in a year or so) a dense cluster of massive stars will become visible at the site of the explosion.

The presence of hydrogen in SN 2006gy is hard to reconcile with the explosion of a  $\gtrsim 40M_{\odot}$  star, because such a star loses its hydrogen-rich envelope several hundreds of thousands of years before the star explodes<sup>6</sup>. Also, the location of the supernova, at a projected distance of about  $1''$  ( $\sim 350$  parsecs, pc) from the nucleus of the host galaxy NGC 1260, is remarkable. A merger between a very massive ( $>100M_{\odot}$ ) hydrogen-depleted star (that already had a core in an advanced phase of helium burning) and a hydrogen-rich main-sequence star of  $10\text{--}40M_{\odot}$  that occurred  $10^4$  to  $10^5$  years before the supernova explosion may explain the unusual brightness of the supernova, the presence of the hydrogen in the interstellar medium surrounding the supernova and the presence of hydrogen in the supernova itself<sup>6,7</sup>.

The existence of young star clusters that are in a state of dynamical core collapse is crucial for the proposed scenario. During core collapse and the subsequent post-core-collapse evolution of the star cluster, a runaway collision product can grow<sup>8</sup>, and even though the star is likely to be much more extended than usual, subsequent bombardment will result in a net increase in mass<sup>9</sup>. Eventually the massive star is expected to prostrate to a black hole of intermediate mass<sup>10,11</sup>. The supernova in which the black hole forms is likely to be unusually bright with some hydrogen in its envelope leftover from the last collision.

The inner few hundred parsecs around the centre of the Milky Way is populated with several bright and dense star clusters, of which the Arches cluster<sup>12</sup> and Quintuplet<sup>13</sup> are the most well known, but many others exist<sup>14–16</sup>. The proximity of the Galactic centre and the depth of the potential well of the bulge causes these clusters to be denser than elsewhere in the Galaxy<sup>17</sup>.

The SB/SB0 host galaxy NGC 1260 appears rather ordinary<sup>18</sup>, although the presence of a dust lane and of H II emission near its centre suggests that a recent burst of star formation occurred near its centre<sup>5,7</sup>. We estimate, taking an intergalactic extinction of  $A \approx 0.43$  magnitudes (mag) (ref. 19) into account, that within the

observed isophotal magnitude  $B_{25} \approx 16''$  ( $\sim 5.6$  kpc)<sup>19</sup> and adopting  $M/L_B \propto L^{0.3}$  (ref. 20), NGC 1260 has a mass of  $M(R = 5.6 \text{ kpc}) \approx 3 \times 10^{10} M_{\odot}$ , where  $R$  is the radius from the centre of NGC 1260. We assume that the mass enclosed within  $R$  is (as is the case for the Milky Way<sup>21</sup>) described with  $M(R) = \mu R^{1.2}$ . For NGC 1260  $\mu \approx 9.5 \times 10^5 M_{\odot}$ , so we can then calculate the lower limit to the tidal radius<sup>22</sup> for a cluster of mass  $m$  in a circular orbit at distance  $R$  from the centre of NGC 1260.

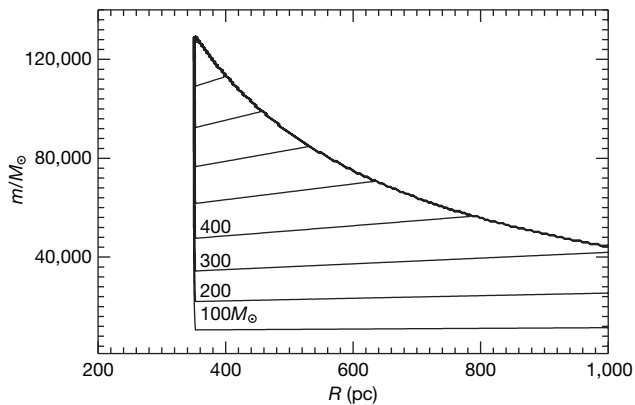
Star clusters that experience core collapse before the most massive stars have left the main sequence can grow a supermassive star via collision runaway<sup>8,23–25</sup>. The mass which can then grow within  $\lesssim 3$  Myr can be estimated using equation (2) of ref. 17. Here we have to make some assumption about the stellar mass function in the cluster, but for clarity, to adopt a mean mass of  $\langle m \rangle = 0.5M_{\odot}$  is sufficient without detailed knowledge of the exact shape of the initial mass function. For a reasonable range of cluster densities and distances from the centre of NGC 1260 we can now calculate the mass that can be grown in the cluster in  $\lesssim 3$  Myr.

In Fig. 1 we present the results of our calculations using a King<sup>26</sup> model with the depth of the central potential expressed by the dimensionless parameter  $W_0 = 8$ , which can produce at most a  $\sim 920M_{\odot}$  star in a collision runaway. For shallower as well as for more concentrated King models the maximum mass for the supermassive star decreases, as well as the mass of the cluster that produces such stars.

The last collision before the supernova must have occurred with a relatively unevolved main-sequence star, and deposited large quantities of hydrogen on the surface of the collision product. By the time of the supernova not all the surface hydrogen of the last collision was blown away, given that about  $1M_{\odot}$  of hydrogen was observed in the supernova<sup>5</sup>. The remainder of the hydrogen deposited on the stellar surface during the last collision was found in the interstellar medium surrounding the supernova, and exceeds some  $0.5\text{--}5M_{\odot}$  (refs 5, 7) but could be as high as  $20\text{--}30M_{\odot}$  (ref. 5). This mass may have come from the strong stellar wind in the last few  $10^4$  years before the supernova, blowing away the hydrogen which was deposited to the stellar surface during the collision. A tentative upper limit for the rate of mass loss of the progenitor star is  $\dot{m} = 5 \times 10^{-4} M_{\odot} \text{ yr}^{-1}$  to  $1.4 \times 10^{-4} M_{\odot} \text{ yr}^{-1}$  (ref. 5). These observed mass-loss rates are consistent with those of detailed evolutionary calculations for stars of  $500\text{--}1,001M_{\odot}$  (refs 27, 28). At this mass-loss rate it takes roughly  $4 \times 10^4$  to  $1.4 \times 10^5$  years to blow  $20M_{\odot}$  in the form of a stellar wind from the surface of the supermassive star. This timescale is of the same order as our estimated average time between collisions of  $\lesssim 7.3 \times 10^4$  years (see Fig. 1).

The luminosity of the supernova explosion in collapsar models is driven by the angular momentum transfer from the critically rotating black hole to its surrounding torus. The available energy reservoir, and thus the supernova brightness, would then be proportional to the

<sup>1</sup>Astronomical Institute 'Anton Pannekoek', <sup>2</sup>Institute for Computer Science, University of Amsterdam, Kruislaan 403, The Netherlands. <sup>3</sup>Kavli Institute for Theoretical Physics, University of California Santa Barbara, California 93106–4030, USA.



**Figure 1 | Mass of the collision runaway star as a function of cluster mass and its distance to the centre of NGC 1260.** The contours, computed for a King model with  $W_0 = 8$ , give the mass of the runaway collision star as a function of the distance to the centre of the galaxy NGC 1260 and the mass of the star cluster. The lowest four contours are labelled by the mass of the supermassive star (in solar masses) with constant increments of  $100M_\odot$  for subsequent curves. A star cluster less massive than about  $6,000M_\odot$  is unable to experience core collapse and produce a collision runaway before it dissolves in the tidal field of the parent galaxy, whereas star clusters in the top right corner are unable to reach core collapse before the most massive stars experience a supernova. The most massive object that can form is  $\sim 920M_\odot$  in a  $m = 1.3 \times 10^5 M_\odot$  cluster. An  $m = 48,000M_\odot$  cluster with  $W_0 = 5$  can produce at most a  $340M_\odot$  supermassive star, whereas a King model with  $W_0 = 11$  can produce a star of at most  $480M_\odot$  in a cluster of  $68,000M_\odot$ . With an average mass increase per collision of  $\sim 20M_\odot$  (refs 8, 23), the supermassive star then has experienced at most some 40 collisions between the moment of gravothermal collapse of the cluster core and the moment that the supermassive star explodes in a supernova. The mean time between collisions for this model is then  $\lesssim 7.3 \times 10^4$  years. For shallower as well as for more concentrated models the time between collisions is larger;  $\lesssim 1.7 \times 10^5$  years for  $W_0 = 5$  and  $\lesssim 1.2 \times 10^5$  years for  $W_0 = 11$ . For a wide range of reasonable cluster parameters it appears likely that a collision runaway ensues and produces a supermassive star.

mass of the black hole<sup>29</sup>. The observed brightness of SN 2006gy would then be consistent with the collapse of an unusually massive star, and the consequent formation of a rather massive ( $\gtrsim 100M_\odot$ ) black hole. We are unaware of detailed simulations of such an unusual supernova to bolster our arguments, but the consequences for the supernova seem to be profound and we encourage further research in this direction.

The amount of hydrogen in the pre-supernova stellar envelope, the amount of hydrogen in the surrounding interstellar medium, the mass-loss rate of the supernova progenitor derived from the observations and the enormous brightness of the supernova all fall within a reasonable uncertainty of the values we derive here on the basis of the collision runaway scenario. We therefore conclude that a collision of a  $\sim 20M_\odot$  main-sequence star with a supermassive star  $\sim 10^5$  years before the supernova could conveniently explain the range of oddities surrounding SN2006gy. We predict that a young ( $\lesssim 5$  Myr), dense and massive ( $10^4 \lesssim m \lesssim 10^5 M_\odot$ ) star cluster is present at the location of the supernova. At the moment the star cluster cannot be seen, but adopting a mass-to-light ratio of  $\sim 0.6$ , which is consistent with the Starburst99 (ref. 30) models for a  $\lesssim 5$ -Myr-old stellar population, the cluster should become noticeable as soon as the supernova fades below an absolute magnitude of about  $-8.2$  mag for a  $10^5 M_\odot$  star cluster, and  $-5.7$  mag for a  $10^4 M_\odot$  star cluster.

The environment in which the collision runaway can be initiated is rather exotic, as the cluster has to be sufficiently massive and dense to warrant dynamical core collapse within a few million years. In star clusters sufficiently massive to grow a massive collision product there are typically between 60 and 600 stars  $> 8M_\odot$ , and consequently only one out of 60–600 type Ib/c or type II supernovae in these clusters will be of this peculiar bright type, as SN2006gy is. If in a nuclear or a

normal starburst ten per cent of all stars are formed in sufficiently dense clusters, one would expect about one out of 600–6,000 supernovae to be of this type.

Received 6 June; accepted 12 September 2007.

- Quimby, R. Supernova 2006gy in NGC 1260. *Central Bureau Electron. Teleg.* **644**, 1 (2006).
- Harutyunyan, A. et al. Supernova 2006gy in NGC 1260. *Central Bureau Electron. Teleg.* **647**, 1 (2006).
- Prieto, J. L., Garnavich, P., Chronister, A. & Connick, P. 2006gy in NGC 1260. *Central Bureau Electron. Teleg.* **648**, 1 (2006).
- Foley, R. J. et al. Supernova 2006gy in NGC 1260. *Central Bureau Electron. Teleg.* **695**, 1 (2006).
- Smith, N. et al. SN 2006gy: discovery of the most luminous supernova ever recorded, powered by the death of an extremely massive star like Eta Carinae. Preprint at (<http://arxiv.org/astro-ph/0002483>) (2006).
- Meynet, G. & Maeder, A. Stellar evolution with rotation. XI. Wolf-Rayet star populations at different metallicities. *Astron. Astrophys.* **429**, 581–598 (2005).
- Ofek, E. O. et al. SN 2006gy: An extremely luminous supernova in the galaxy NGC 1260. *Astrophys. J. Lett.* **659**, L13–L16 (2007).
- Portegies Zwart, S. F., Makino, J., McMillan, S. L. W. & Hut, P. Star cluster ecology. III. Runaway collisions in young compact star clusters. *Astron. Astrophys.* **348**, 117–126 (1999).
- Suzuki, T. K. et al. Evolution of collisionally merged massive stars. Preprint at (<http://arxiv.org/astro-ph/0703290>) (2007).
- Fryer, C. L. & Kalogera, V. Theoretical black hole mass distributions. *Astrophys. J.* **554**, 548–560 (2001).
- Portegies Zwart, S. F., Baumgardt, H., Hut, P., Makino, J. & McMillan, S. L. W. Formation of massive black holes through runaway collisions in dense young star clusters. *Nature* **428**, 724–726 (2004).
- Cotera, A. S. et al. The discovery of hot stars near the galactic center thermal radio filaments. *Astrophys. J.* **461**, 750–761 (1996).
- Cotera, A. S. et al. A new cluster of hot stars near the galactic center. *Bull. Am. Astron. Soc.* **1262**–1263 (1992).
- Dutra, C. M. & Bica, E. New star clusters projected close to the galactic centre. *Astron. Astrophys.* **359**, L9–L12 (2000).
- Dutra, C. M. et al. NTT infrared imaging of star cluster candidates towards the central parts of the Galaxy. *Astron. Astrophys.* **408**, 127–134 (2003).
- Maillard, J. P., Paumard, T., Stoklov, S. R. & Rigaut, F. The nature of the Galactic Center source IRS 13 revealed by high spatial resolution in the infrared. *Astron. Astrophys.* **423**, 155–167 (2004).
- Portegies Zwart, S. F. et al. The ecology of star clusters and intermediate-mass black holes in the Galactic Bulge. *Astrophys. J.* **641**, 319–326 (2006).
- Brunzendorf, J. & Meusinger, H. The galaxy cluster Abell 426 (Perseus). A catalogue of 660 galaxy positions, isophotal magnitudes and morphological types. *Astron. Astrophys. (Suppl.)* **139**, 141–161 (1999).
- Wegner, G. et al. Redshift-distance survey of early-type galaxies: spectroscopic data. *Astron. J.* **126**, 2268–2280 (2003).
- Salucci, P., Ashman, K. M. & Persic, M. The dark matter content of spiral galaxies. *Astrophys. J.* **379**, 89–93 (1991).
- Sanders, R. H. & Lowinger, T. The distribution of mass in the galactic nucleus. *Astron. J.* **77**, 292–297 (1972).
- Binney, J. & Tremaine, S. *Galactic Dynamics* 747 (Princeton Univ. Press, Princeton, New Jersey, 1987).
- Portegies Zwart, S. F. & McMillan, S. L. W. The runaway growth of intermediate-mass black holes in dense star clusters. *Astrophys. J.* **576**, 899–907 (2002).
- Gürkan, M. A., Freitag, M. & Rasio, F. A. Formation of massive black holes in dense star clusters. I. Mass segregation and core collapse. *Astrophys. J.* **604**, 632–652 (2004).
- Freitag, M., Gürkan, M. A. & Rasio, F. A. Runaway collisions in young star clusters—II. Numerical results. *Mon. Not. R. Astron. Soc.* **368**, 141–161 (2006).
- King, I. R. The structure of star clusters. III. Some simple dynamical models. *Astron. J.* **71**, 64–75 (1966).
- Yungelson, L. R. Evolution of Ultra-Massive Stars and its Implications. Presented at 26th General Assembly of the IAU in Joint Discussion 5. “Calibrating the Top of the Stellar M-L Relation” (Prague, 2006).
- Yungelson, L. R., van den Heuvel, E. P. J., Vink, J. S., Portegies Zwart, S. F. & de Koter, A. On the evolution and fate of supermassive stars. *Astron. Astrophys.* (in the press); preprint available at (<http://arxiv.org/abs/0710.1181v1>) (2007).
- Lee, H. K., Wijers, R. A. M. J. & Brown, G. E. The Blandford-Znajek process as a central engine for a gamma-ray burst. *Phys. Rep.* **325**, 83–114 (2000).
- Leitherer, C. et al. Starburst99: synthesis models for galaxies with active star formation. *Astrophys. J. (Suppl.)* **123**, 3–40 (1999).

**Acknowledgements** This work was supported by the NWO, the Netherlands Research School for Astronomy (NOVA) and by a grant from the National Science Foundation to the Kavli Institute for Theoretical Physics in Santa Barbara.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to S.F.P.Z. (spz@science.uva.nl) or E.P.J.v.d.H. (edvdh@science.uva.nl).



## LETTERS

# Pulsational pair instability as an explanation for the most luminous supernovae

S. E. Woosley<sup>1</sup>, S. Blinnikov<sup>1,2,3</sup> & Alexander Heger<sup>1,4</sup>

The extremely luminous supernova SN 2006gy (ref. 1) challenges the traditional view that the collapse of a stellar core is the only mechanism by which a massive star makes a supernova, because it seems too luminous by more than a factor of ten. Here we report that the brightest supernovae in the modern Universe arise from collisions between shells of matter ejected by massive stars that undergo an interior instability arising from the production of electron–positron pairs<sup>2</sup>. This ‘pair instability’ leads to explosive burning that is insufficient to unbind the star, but ejects many solar masses of the envelope. After the first explosion, the remaining core contracts and searches for a stable burning state. When the next explosion occurs, several solar masses of material are again ejected, which collide with the earlier ejecta. This collision can radiate  $10^{50}$  erg of light, about a factor of ten more than an ordinary supernova. Our model is in good agreement with the observed light curve for SN 2006gy and also shows that some massive stars can produce more than one supernova-like outburst.

The life of a star is determined by the mass, composition and rotation rate with which it is born. Most important to its death is the mass, at the end of the star’s life, of its core of helium and heavy elements, usually called the ‘helium core’ (Table 1). That mass, which determines the explosion mechanism for the supernova the star makes and its nucleosynthesis, is in turn sensitive to how much mass the star lost along the way. For currently favoured mass loss rates and solar composition, stars with initial masses over  $\sim 40$  times that of the Sun lose all their hydrogen envelope and part of their helium core as well. Current calculations suggest that the maximum helium core at death is only about 15 solar masses<sup>3,4</sup>, and no modern star, at least in our Galaxy, would encounter the pair instability. The rate at which the most massive stars lose mass is quite uncertain<sup>5–7</sup> though, and depends upon their metal content<sup>8,9</sup>. Much more massive helium cores could have been common for stars born in the distant past and perhaps, occasionally, even today.

Among these most massive stars, particularly poorly explored are ‘pulsational pair-instability supernovae’<sup>2,10,11</sup>, which might occur at the deaths of main-sequence stars in the mass range 95 to 130 solar masses (Table 1). The pair instability is encountered when, late in the

star’s life, a large amount of thermal energy goes into making the masses of an increasing abundance of electron–positron pairs rather than providing pressure. Rapid contraction occurs, followed by a thermonuclear explosion<sup>2,12,13</sup>. But in the pulsational case, the energy released by the explosive burning is inadequate to unbind the entire star. It suffices, however, violently to eject many solar masses of surface material, including all that is left of the hydrogen envelope, in a series of giant ‘pulses’. The typical binding energy for the hydrogen envelope of such massive stars is only  $\sim 0.1$  to  $1 \times 10^{49}$  erg, whereas the energy of a pulse is  $\sim 1$ – $100 \times 10^{49}$  erg (Supplementary Table 1), so the envelope is easily ejected in the first pulse. After each pulse, the remaining core contracts, radiates neutrinos and light, and searches again for a stable burning state. The time required for this contraction is sensitive to the strength of the pulse and how close the star came to becoming unbound. If the temperature after the first pulse is less than about  $9 \times 10^8$  K, neutrino losses are inefficient and it may be decades before the star starts burning again. If the core is much hotter, it may only take days.

If the remaining helium core is still over 40 solar masses, with the exact threshold depending upon the entropy lost to neutrinos during the interpulse period, the star encounters the instability again, and ejects another several solar masses. Later ejections have lower mass, because the envelope was expelled in the first pulse, but have higher energy. They quickly catch up to the first shell, which by this time is at  $10^{15}$ – $10^{16}$  cm, where the collision dissipates most of their relative kinetic energy as radiation (Fig. 1). Because of the large radius for the collision, adiabatic losses from expansion are roughly two orders of magnitude less than in a common type-II supernova. That is, a collision involving only  $10^{50}$  erg of kinetic energy can radiate as much as  $10^{50}$  erg of light, more than ten times an ordinary supernova.

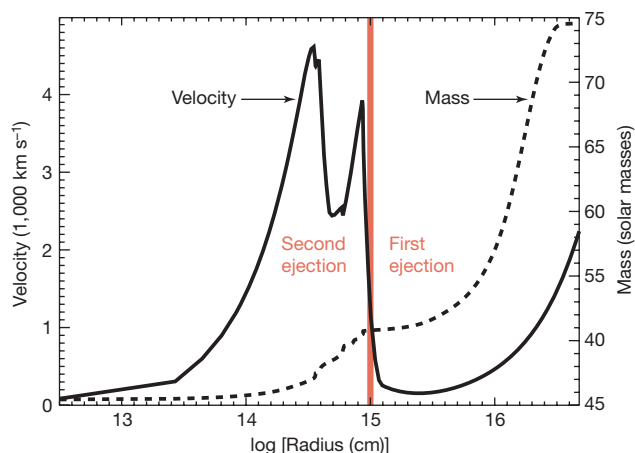
To illustrate these general ideas, consider the evolution of a star of 110 solar masses and solar composition. Its evolution is calculated using the Kepler code<sup>3,14</sup> with mass loss included at a fraction of the standard value for solar metallicity stars<sup>15,16</sup>, 50% on the main sequence, 10% as a helium-burning red giant. Our 110-solar-mass main-sequence star then ends its life with a total mass of 74.6 solar masses and a helium core of 49.9 solar masses, well within the

**Table 1 | Final evolution of stars of different initial mass<sup>3</sup>**

Mass at birth (solar masses)	Helium core mass (solar masses)	Compact remnant	Event
10–95	2–40	Neutron star, black hole	Ordinary supernova
95–130	40–60	Neutron star, black hole	Pulsational pair-instability supernova
130–260	60–137	Explosion, no remnant	Pair-instability supernova
>260	>137	Black hole	?

Column 1 gives the total mass of the (non-rotating) star when it is born. If the outer layers of hydrogen and helium are not entirely lost along the way, the second column gives the mass of the core of helium and heavier elements inside the star when it dies. Columns 3 and 4 then describe how the star dies and what sort of remnant it leaves behind. Without rotation, helium cores over 137 solar masses simply disappear into a black hole. With rotation, their evolution is uncertain. The pair instability occurs after carbon burning when the centre of the star encounters thermodynamic conditions where a large fraction of the internal energy is stored in the rest masses of electron–positron pairs. The loss of pressure renders the star briefly unstable against collapse, nuclear burning and explosion.

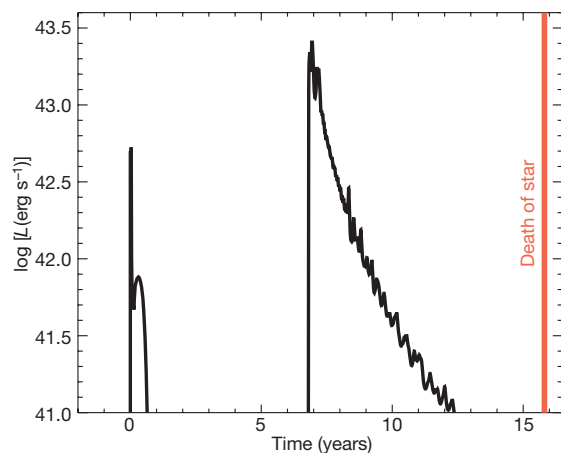
<sup>1</sup>Department of Astronomy and Astrophysics, UCSC, Santa Cruz, California 95064, USA. <sup>2</sup>ITEP, 117218 Moscow, Russia. <sup>3</sup>Max Planck Institut für Astrophysik, Garching, D-85741, Germany. <sup>4</sup>Theoretical Astrophysics Group, T-6, MS B227, Los Alamos National Laboratory, Los Alamos, New Mexico 87544, USA.



**Figure 1 | Velocity structure following the second eruption of a 110-solar-mass pulsational pair-instability supernova.** The velocity and enclosed mass are plotted against the log of the radius. The velocity discontinuity at  $10^{15}$  cm shows where fast-moving ejecta from the second outburst are starting to collide with the slower-moving material ejected in the first pulse. Hydrogen-rich and helium-rich material immediately above this shock is moving at less than  $200 \text{ km s}^{-1}$  and will give rise to narrow lines in the spectrum of the emission, as was seen in SN 2006gy<sup>1</sup>. Most of the kinetic energy of the second ejection will be dissipated within  $10^{16}$  cm. This particular configuration resulted from a star initially with 110 solar masses that had 74.6 solar masses left when it began exploding (see text).

pulsational domain. The pre-supernova star is a red supergiant with radius  $1.1 \times 10^{14}$  cm and luminosity  $9.2 \times 10^{39} \text{ erg s}^{-1}$ . Its outer 24 solar masses of low-density envelope are only bound by  $9.0 \times 10^{48}$  erg.

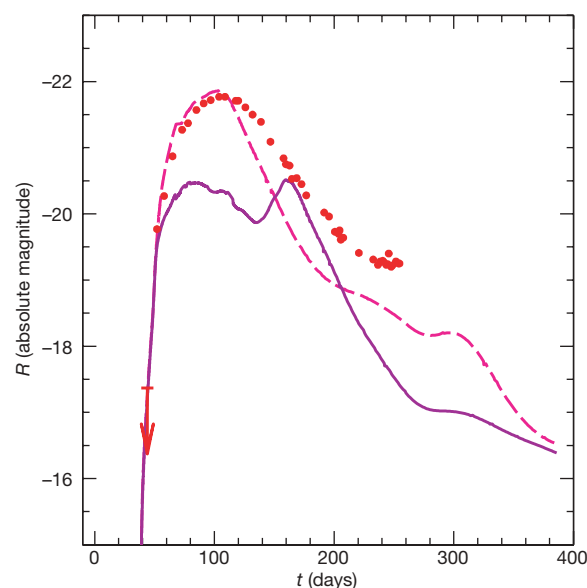
After burning helium and carbon, when the temperature exceeds  $10^9$  K, this star first encounters the pair instability. The helium core collapses rapidly to a maximum central temperature of  $3.04 \times 10^9$  K and density  $1.50 \times 10^6 \text{ gm cm}^{-3}$ , far hotter than the usual  $2.0 \times 10^9$  K at which oxygen burns stably in a massive star. So the star violently explodes, burning 1.49 solar masses of oxygen and 1.55 solar masses of carbon and releasing  $1.4 \times 10^{51}$  erg (Supplementary Figs 1 and 2).



**Figure 2 | Cumulative light curve for the 110-solar-mass model.** Three events characterize the final years of the star's life. The first major eruption ejects about 25 solar masses of hydrogen–helium envelope and makes a supernova with luminosity  $\sim 6 \times 10^{41} \text{ erg s}^{-1}$  lasting 200 days (see Supplementary Fig. 5). Shock breakout produces the brief bright ultraviolet transient at the onset of this first light curve, while the plateau is due to hydrogen recombination. Then 6.9 years later a second eruption produces a brilliant event as the fast-moving ejecta collide with the debris of the first supernova (Fig. 1 and 3). And 9 years after that, the star forms a 2.2-solar-mass iron core that collapses to a rapidly rotating neutron star or black hole. A third bright event, possibly a gamma-ray burst, might then occur.

Most of this energy goes into expanding the star. About 10%, however, goes into driving off 24.5 solar masses of envelope and core (mostly helium and some hydrogen) with a terminal speed of 100 to  $1,000 \text{ km s}^{-1}$  (Fig. 1). This envelope ejection gives the first supernova-like display with a luminosity  $\sim 4 \times 10^{41} \text{ erg s}^{-1}$  for 200 days (Fig. 2 and Supplementary Fig. 5).

What is left behind is a 50.7-solar-mass remnant, slightly larger than the original helium core mass, that once again radiates neutrinos, contracts and grows hotter. Then 6.8 years later, it encounters the pair instability a second time. This time the pulse is stronger, and  $6.0 \times 10^{50}$  erg is shared by a smaller ejected mass of 5.1 solar masses. The collision of this high-velocity shell with the larger mass ejected earlier (Fig. 1) produces a brilliant light curve<sup>17</sup> calculated here using the radiation-hydrodynamics code Stella<sup>18</sup> (Fig. 3). Stella uses multi-energy groups to compute the coupling of radiation transfer to the gas dynamics and produces multi-colour and bolometric light curves. Previously Stella was used successfully to resolve a very thin shell and radiative shock in the case of SN 1994W and gave multi-colour fluxes in good agreement with observations<sup>19</sup>. The good agreement with SN 2006gy<sup>1</sup>, (Fig. 3 and Supplementary Figs 7–9) is suggestive of a light curve generated by collisions between solar masses of material as refs 1 and 20 have also proposed. Other models for SN 2006gy based upon traditional pair-instability supernovae can be very bright<sup>12,21</sup>, but require a large mass of  $^{56}\text{Ni}$  and are difficult



**Figure 3 | Absolute R-band magnitudes resulting from the strong second explosion of the 110-solar-mass model.** The time axis has been adjusted so as to give the best agreement with observations of SN 2006gy<sup>1</sup>, plotted as the red data points, and the model results have been smoothed using a numerical averaging over 30-day intervals. Multi-dimensional calculations of similar models<sup>21</sup> suggest that instabilities in the thin dense shell, where the radiation originates, will result in the formation of a mixed layer with relative thickness  $\Delta R/R \approx 0.1$ – $0.15$ . The predictions of our one-dimensional model (where the radius is proportional to the time) should thus be blurred by  $\Delta t \approx 30$  days for a total light-curve width of about 200 days. An R-band extinction of 1.68 magnitudes is assumed for the supernova<sup>1</sup>. Two curves are shown, one for the nominal model discussed in the text, and a second where the velocity of all the ejecta—pulses 1 and 2—has been multiplied by two (hence an artificial increase in the explosion energy from  $7.2 \times 10^{50}$  erg to  $2.9 \times 10^{51}$  erg). The large variations apparent in the fainter model are absent in the brighter one because the photosphere at peak light in the more energetic model has not receded to near the shock. The actual explosion energy and mass ejected are sensitive to the initial mass of the star, its uncertain mass loss (that is, the mass of the remaining hydrogen envelope when the star dies), and details of the cooling between pulses. Other light curves, without smoothing and with variable explosion energy and density, are given in Supplementary Figs 7–9.

to reconcile with the narrow width of the observed light curve<sup>22,23</sup> and the narrow spectral features due to hydrogen. They also require exceptionally massive progenitor stars.

The photospheric structure of these collisionally dominated supernovae is novel. Early on, the collision occurs at such high density and small radius that the shock is preceded by an optically thick photosphere (Supplementary Figs 10 and 11). Matter inside this photosphere is ionized and optically thick. The emission is nearly blackbody and no X-ray or radio emission is produced. The mass in the second eruption quickly becomes concentrated in very thin shells. Later, the large velocity shear bounding these shells keeps the opacity in Doppler-broadened lines from becoming too small, and the emission continues to be predominantly in near-optical bands. The large column depth probably keeps any appreciable X-rays that are produced from escaping until after the optical display is over. This is consistent with the low level of X-rays detected from the supernova<sup>20</sup>.

Nine years later, the 110-solar-mass model finishes a final phase of contraction and gently starts silicon burning at its centre, making an iron core that collapses (Fig. 2 and Supplementary Fig. 4). A 95-solar-mass star that evolved similarly, but with mild rotation (equatorial speed  $100 \text{ km s}^{-1}$  on the main sequence) and magnetic torques<sup>24</sup>, produces a similar helium-core mass, but has sufficient angular momentum in its iron core to make a neutron star with period 2 ms. This is sufficiently rapid rotation to form a magnetar<sup>25</sup>, or, within uncertainties in the angular momentum transport model, a collapsar<sup>26</sup>. Thus the final death of the star might generate a gamma-ray burst<sup>27,28</sup>, but one that is embedded in many solar masses of circumstellar material. The optical light curve from such an event could also be very bright and might be an alternative explanation for SN 2006gy.

Indeed, as Supplementary Table 1 shows, the pulsational pair-instability mechanism can energize a variety of explosive phenomena with characteristic timescales ranging from days to centuries. We have focused here on the brightest of these events, but if the energy of the ejected shells is low and if the core of the star eventually collapses to a slowly rotating black hole, we might observe “supernova impostors”<sup>29</sup> and nothing else. If the star lost its envelope, but retained a supercritical helium core mass, it might form a repeating type Ib supernova<sup>30</sup>.

Received 4 June; accepted 21 September 2007.

- Smith, N. *et al.* SN 2006gy: discovery of the most luminous supernova ever recorded, powered by the death of an extremely massive star like Eta Carinae. *Astrophys. J.* **666**, 1116–1128 (2007).
- Barkat, Z., Rakavy, G. & Sack, N. Dynamics of supernova explosions resulting from pair formation. *Phys. Rev. Lett.* **18**, 379–381 (1967).
- Woosley, S. E., Heger, A. & Weaver, T. A. The evolution and explosion of massive stars. *Rev. Mod. Phys.* **74**, 1015–1071 (2002).
- Figer, D. F. An upper limit to the masses of stars. *Nature* **434**, 192–194 (2005).
- Vink, J. S. & de Koter, A. Predictions of variable mass loss for Luminous Blue Variables. *Astron. Astrophys.* **393**, 543–553 (2002).
- Smith, N. & Owocki, S. P. On the role of continuum-driven eruptions in the evolution of very massive stars and population III stars. *Astrophys. J. Lett.* **645**, L45–L48 (2006).
- Smith, N. Eruptive mass loss in very massive stars and population III stars. In *Massive Stars: From Pop III and GRBs to the Milky Way* (in the press); preprint at (<http://arxiv.org/abs/astro-ph/0607457v2>) (2006).
- Vink, J. S. & de Koter, A. On the metallicity dependence of Wolf-Rayet wind. *Astron. Astrophys.* **442**, 587–596 (2005).
- Kudritzki, R. P. Line-driven winds, ionizing fluxes, and ultraviolet spectra of hot stars at extremely low metallicity. I. Very massive O stars. *Astrophys. J.* **577**, 389–408 (2002).

- Woosley, S. E. & Weaver, T. A. In *Radiation Hydrodynamics in Stars and Compact Objects* (eds Mihalas, D. & Winkler, K.-H. A.) *Lecture Notes Phys.* **255**, 91–120 (1986).
- Heger, A. & Woosley, S. E. The nucleosynthetic signature of population III. *Astrophys. J.* **567**, 532–543 (2002).
- Woosley, S. E. & Weaver, T. A. In *Supernovae: A Survey of Current Research* (eds Rees, M. J. & Stoneham, R. J.) *NATO Advanced Study Inst. Ser.* **90**, 79–99 (1982).
- Bond, J. R., Arnett, W. D. & Carr, B. The evolution and fate of very massive objects. *Astrophys. J.* **280**, 825–847 (1984).
- Weaver, T. A., Zimmerman, G. B. & Woosley, S. E. Presupernova evolution of massive stars. *Astrophys. J.* **225**, 1021–1029 (1978).
- de Jager, C., Nieuwenhuijzen, H. & van der Hucht, K. A. Mass loss rates in the Hertzsprung–Russell diagram. *Astron. Astrophys.* **72** (Suppl.), 259–289 (1988).
- Nieuwenhuijzen, H. & de Jager, C. Parametrization of stellar rates of mass loss as functions of the fundamental stellar parameters M, L, and R. *Astron. Astrophys.* **231**, 134–136 (1990).
- Grasberg, E. K. & Nadyozhin, D. K. Type-II supernovae—two successive explosions. *Sov. Astron. Lett.* **12**, 68–70 (1986).
- Blinnikov, S. I. *et al.* Theoretical light curves for deflagration models of type Ia supernova. *Astron. Astrophys.* **453**, 229–240 (2006).
- Chugai, N. N. *et al.* Type II supernova 1994W: evidence for the explosive ejection of a circumstellar envelope. *Mon. Not. R. Astron. Soc.* **352**, 1213–1231 (2004).
- Ofek, E. O. *et al.* SN 2006gy: An extremely luminous supernova in the galaxy NGC 1260. *Astrophys. J. Lett.* **659**, L13–L16 (2007).
- Scannapieco, E., Madau, P., Woosley, S. E., Heger, A. & Ferrara, A. The detectability of pair-production supernovae at  $z < 6$ . *Astrophys. J.* **633**, 1031–1041 (2005).
- Nomoto, K., Tominaga, N., Tanaka, M., Maeda, K. & Umeda, H. Nucleosynthesis in core-collapse supernovae and GRB-metal-poor star connection. In *Supernova 1987A: 20 Years After: Supernovae and Gamma-Ray Bursters* (eds Immler, S., Weiler, K. & McCray, R.) (American Institute of Physics, in the press); preprint at (<http://arxiv.org/abs/0707.2187>) (2007).
- Umeda, H. & Nomoto, K. How much  $^{56}\text{Ni}$  can be produced in core-collapse supernovae? Evolution and explosion of 30–100 solar mass stars. *Astrophys. J.* (submitted); preprint at (<http://arxiv.org/astro-ph/07072598>) (2007).
- Heger, A., Woosley, S. E. & Spruit, H. Presupernova evolution of differentially rotating massive stars including magnetic fields. *Astrophys. J.* **626**, 350–363 (2005).
- Duncan, R. C. & Thompson, C. Formation of very strongly magnetized neutron stars—Implications for gamma-ray bursts. *Astrophys. J.* **392**, L9–L13 (1992).
- Woosley, S. E. Gamma-ray bursts from stellar mass accretion disks around black holes. *Astrophys. J.* **405**, 273–277 (1993).
- Usov, V. V. Millisecond pulsars with extremely strong magnetic fields as a cosmological source of gamma-ray bursts. *Nature* **357**, 472–474 (1992).
- MacFadyen, A. I. & Woosley, S. E. Collapsars: gamma-ray bursts and explosions in “failed supernovae”. *Astrophys. J.* **524**, 262–289 (1999).
- Van Dyk, S. D., Filippenko, A. V., Chornock, R., Li, W. & Challis, P. M. Supernova 1954J (variable 12) in NGC 2403 unmasked. *Publ. Astron. Soc. Pacif.* **117**, 553–562 (2005).
- Foley, R. J. *et al.* 2006jc: a Wolf-Rayet star exploding in a dense He-rich circumstellar medium. *Astrophys. J.* **657**, L105–L108 (2007).
- Chevalier, R. A. & Blondin, J. M. Hydrodynamic instabilities in supernova remnants: early radiative cooling. *Astrophys. J.* **444**, 312–317 (1995).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported by the Scientific Discovery through Advanced Computing (SciDAC) Program of the US Department of Energy, by NASA, and by the Russian Foundation for Basic Research and Science Schools. At Los Alamos, this work was carried out under the auspices of the National Nuclear Security Administration of the US Department of Energy.

**Author Contributions** S.E.W. and A.H. proposed that the light curves of pulsational pair-instability supernovae might have a large range in luminosity including exceptionally brilliant supernovae. They carried out the calculations of stellar evolution and explosion. S.B. provided expertise in the physics of supernovae with circumstellar interaction and calculated all the light curves from the models except those done with Kepler.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to S.E.W. ([woosley@ucolick.org](mailto:woosley@ucolick.org)).



# Entanglement-free Heisenberg-limited phase estimation

B. L. Higgins<sup>1</sup>, D. W. Berry<sup>2</sup>, S. D. Bartlett<sup>3</sup>, H. M. Wiseman<sup>1,4</sup> & G. J. Pryde<sup>1</sup>

Measurement underpins all quantitative science. A key example is the measurement of optical phase, used in length metrology and many other applications. Advances in precision measurement have consistently led to important scientific discoveries. At the fundamental level, measurement precision is limited by the number  $N$  of quantum resources (such as photons) that are used. Standard measurement schemes, using each resource independently, lead to a phase uncertainty that scales as  $1/\sqrt{N}$ —known as the standard quantum limit. However, it has long been conjectured<sup>1,2</sup> that it should be possible to achieve a precision limited only by the Heisenberg uncertainty principle, dramatically improving the scaling to  $1/N$  (ref. 3). It is commonly thought that achieving this improvement requires the use of exotic quantum entangled states, such as the NOON state<sup>4,5</sup>. These states are extremely difficult to generate. Measurement schemes with counted photons or ions have been performed with  $N \leq 6$  (refs 6–15), but few have surpassed the standard quantum limit<sup>12,14</sup> and none have shown Heisenberg-limited scaling. Here we demonstrate experimentally a Heisenberg-limited phase estimation procedure. We replace entangled input states with multiple applications of the phase shift on unentangled single-photon states. We generalize Kitaev's phase estimation algorithm<sup>16</sup> using adaptive measurement theory<sup>17–20</sup> to achieve a standard deviation scaling at the Heisenberg limit. For the largest number of resources used ( $N = 378$ ), we estimate an unknown phase with a variance more than 10 dB below the standard quantum limit; achieving this variance would require more than 4,000 resources using standard interferometry. Our results represent a drastic reduction in the complexity of achieving quantum-enhanced measurement precision.

Phase estimation is a ubiquitous measurement primitive, used for precision measurement of length, displacement, speed, optical properties, and much more. Recent work in quantum interferometry has focused on  $n$ -photon NOON states<sup>5–12,21</sup>,  $(|n\rangle|0\rangle + |0\rangle|n\rangle)/\sqrt{2}$ , expressed in terms of number states of the two arms of the interferometer. With this state, an improved phase sensitivity results from a decrease in the phase period from  $2\pi$  to  $2\pi/n$ . We achieve improved phase sensitivity more simply using an insight from quantum computing. We apply Kitaev's phase estimation algorithm<sup>16,22</sup> to quantum interferometry, wherein the entangled input state is replaced by multiple passes through the phase shift. The idea of using multi-pass protocols to gain a quantum advantage was proposed for the problem of aligning spatial reference frames<sup>23</sup>, and further developed in relation to clock synchronization<sup>24</sup> and phase estimation<sup>25,26</sup>.

The conceptual circuit for Kitaev's phase estimation algorithm is shown in Fig. 1a. The algorithm yields, with  $K + 1$  bits of precision, an estimate  $\phi_{\text{est}}$  of a classical phase parameter  $\phi$ , where  $e^{i\phi}$  is an eigenvalue of a unitary operator  $U$ . It requires us to apply  $K + 1$

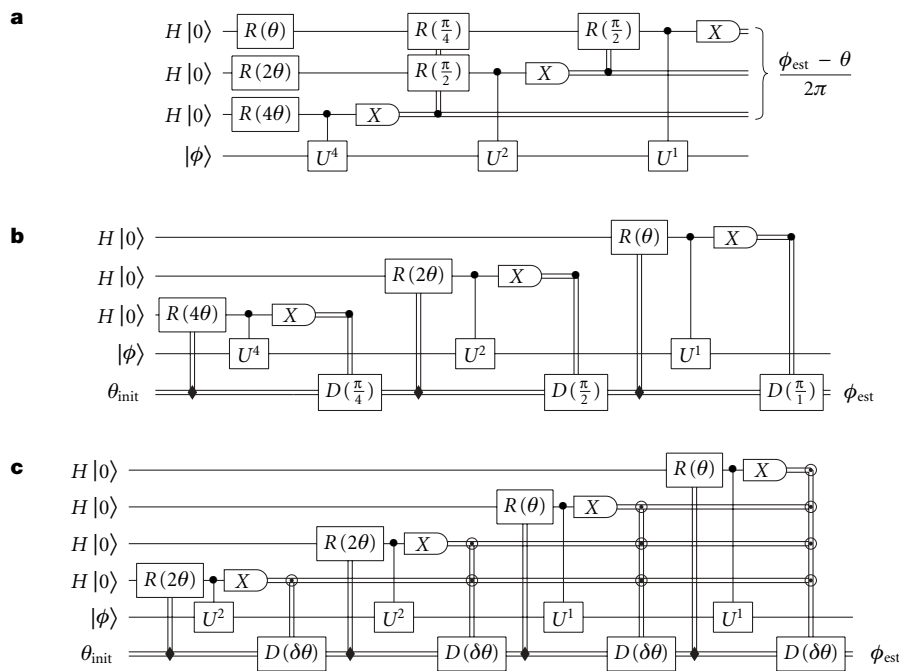
unitaries,  $U^p$ , with  $p = 2^K, 2^{K-1}, \dots, 1$ , each controlled by a different qubit. Each qubit is prepared in the state  $H|0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ , and the control induces a phase shift  $e^{ip\phi}$  on the  $|1\rangle$  component. The qubits are measured sequentially in the  $\sigma_x$  basis ( $X$ ), and the results control additional phase shifts, indicated by  $R(\alpha) \equiv \exp(i\alpha|0\rangle\langle 0|)$ , on subsequent qubits. This enables the inverse quantum Fourier transform to be performed without entangling gates<sup>27</sup>. With a random phase  $\theta$  on the qubits, as shown in Fig. 1a, the measurement results on the qubits are the binary digits of  $(\phi_{\text{est}} - \theta)/2\pi$ ; this ensures that the accuracy of the estimate is independent of the value of  $\phi$ .

Alternatively, this independence could be obtained by using a second classical 'feedback' phase  $\theta$ , as in Fig. 1b, which also eliminates the need for many of the gates in Fig. 1a. This is a classical real-valued parameter whose value is adjusted by  $\pi/p$ , indicated by the symbol  $D(\pi/p)$ , controlled by the results of measurements. The value of  $\theta$  determines (as indicated by the diamond-shaped control symbol in Fig. 1b) phase-shifts  $R(p\theta)$  on the qubits. Applying this to interferometry, we can measure an unknown optical phase  $\phi$  using dual-rail photonic qubits<sup>22</sup>. Here the operator  $U$  induces a relative phase shift  $\phi$  each time the beam path (in one arm of the interferometer) passes through the unknown optical phase  $\phi$ . The additional phase shifts (determined by  $\theta$ ) can be implemented using a single-pass controllable phase in the other arm.

If a fixed probability of error in  $\phi_{\text{est}}$  is allowed (that is, if the uncertainty is quantified by a confidence interval), then the uncertainty of Kitaev's phase estimation scales as  $2^{-K}$  (ref. 22). Because the number of control photons is  $N_{\text{phot}} = K + 1$ , this scaling implies an exponential decrease in the phase uncertainty with increasing resources—apparently violating the Heisenberg uncertainty principle. The correct analysis, however, is as follows. Although the cost of implementing  $U^p$  can be assumed to be essentially independent of  $p$  in the context of quantum computation, in interferometry it requires  $p$  applications of the phase shift, and should thus be counted as requiring  $p$  resources<sup>25</sup>. Using this definition, the total number of resources used is  $N = 2^{K+1} - 1$ . Then for  $N \gg 1$ , the uncertainty scales as  $1/N$ , as in the Heisenberg limit. We note that this quantification of resources in terms of the number of applications of the phase shift is the relevant one for phase estimation of sensitive (for example, biological) samples, wherein the goal is to pass as little light through the sample as is necessary.

On the other hand, if  $\Delta\phi_{\text{est}}$  is taken to be the standard deviation—the usual measure of uncertainty—then Kitaev's algorithm does not scale as  $1/N$ . Rather, we have shown analytically that it asymptotes as  $\sqrt{2}/\sqrt{N}$ , the same scaling as the standard quantum limit (SQL)—see also ref. 21. The broad wings of the distribution of phase estimates are not due to any deficiency in the estimation procedure—the quantum Fourier transform is optimal—but rather are a consequence of the sequence of phase shifts on the photons,  $2^K\phi, 2^{K-1}\phi, \dots, \phi$ .

<sup>1</sup>Centre for Quantum Dynamics, Griffith University, Brisbane 4111, Australia. <sup>2</sup>Centre for Quantum Computer Technology, Macquarie University, Sydney 2109, Australia. <sup>3</sup>School of Physics, University of Sydney, Sydney 2006, Australia. <sup>4</sup>Centre for Quantum Computer Technology, Griffith University, Brisbane 4111, Australia.



**Figure 1 | Quantum circuit diagrams of Kitaev's phase estimation algorithm and our generalization.** **a**, Kitaev's algorithm<sup>16</sup> with the inverse quantum Fourier transform implemented with measurement and classical feedback<sup>27</sup> and a random initial phase estimate  $\theta$ . In general,  $K + 1$  qubits yield  $K + 1$  binary digits of precision; here  $K = 2$ . **b**, As in **a**, but here we

implement  $\theta$  (now called  $\theta_{\text{init}}$ ) and the feedback operations by coupling the qubits to a common element, the 'feedback phase'  $\theta$  (the lowest rail). **c**, Generalization of the circuit to include  $M \geq 1$  qubits for each binary digit; here  $K = 1$  and  $M = 2$ . For details on circuit elements, see text.

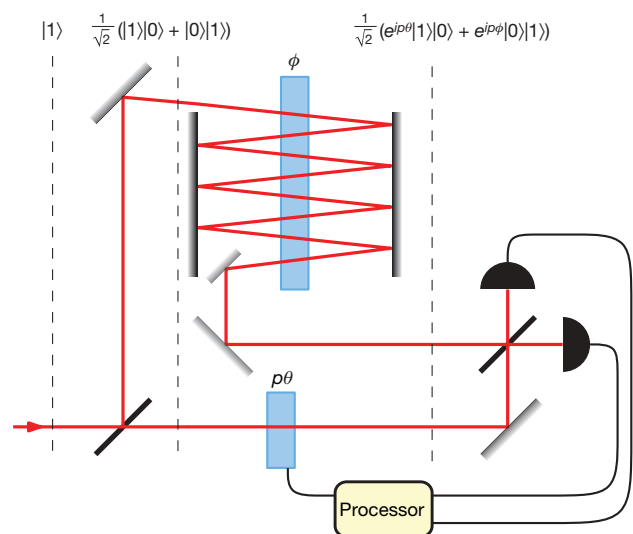
A key idea to address this problem is to employ  $M$  copies of the control photon at each phase shift<sup>23–25</sup>. For  $M > 1$  one cannot perform an exact quantum Fourier transform using only single-photon operations. However, one can perform it approximately using the adaptive phase estimation scheme of ref. 18, as shown in Fig. 1c for  $M = 2$ . Here the feedback phase is adjusted by an amount  $\delta\theta$ , controlled by all previous measurement results via a Bayesian algorithm. This general multi-bit conditioning is represented by the circled-dot symbol in Fig. 1c. For the final adjustment,  $\delta\theta = \phi_{\text{est}} - \theta$ , so the value of  $\theta$  that is read out is equal to  $\phi_{\text{est}}$ , the best estimate of the phase given all the data, as in Fig. 1b. Because the inverse quantum Fourier transform is not performed exactly, and because the sequence of phase shifts is not exactly equivalent to the optimal state of ref. 18, we do not expect  $\phi_{\text{est}}$  to have an uncertainty precisely at the Heisenberg limit  $\Delta\phi_{\text{HL}} = \tan[\pi/(N + 2)] \approx \pi/N$  for  $N \gg 1$  (refs 18, 28). Nevertheless, our algorithm allows estimation with an uncertainty only a constant factor larger than this ultimate limit, for  $M \geq 4$ . For instance, we find by numerical simulation that for  $M = 6$ ,  $\Delta\phi_{\text{est}} \approx 1.56\pi/N$  for  $N \gg 1$ .

A conceptual implementation of this generalization of Kitaev's algorithm is shown in Fig. 2. It works as follows: a photon is converted to the state  $\frac{1}{\sqrt{2}}(|1\rangle|0\rangle + |0\rangle|1\rangle)$  by the first beam splitter. After passing  $p = 2^K$  times through the phase shift  $\phi$ , the state evolves to  $\frac{1}{\sqrt{2}}(e^{ip\theta}|1\rangle|0\rangle + e^{ip\phi}|0\rangle|1\rangle)$ . The photon is then detected after the modes are recombined on the second beam splitter. The result is used to update the probability distribution  $P(\phi)$ , which represents knowledge about  $\phi$ .

This process is repeated  $M$  times, so that  $M$  independent photons go through  $2^K$  passes in sequence. Quantifying a resource as a single pass of a photon through the phase shift, each photon in this stage corresponds to  $2^K$  resources. Following these  $M$  photons, another  $M$  photons undergo the same process at  $p = 2^{K-1}$  passes, and so on for  $p = 2^{K-2}, \dots, 2^0$ . Thus a total of  $M(K + 1)$  photons and  $N = M(2^{K+1} - 1)$  resources are used. The value of the feedback phase  $\theta$  is random for the first photon only. Thereafter it is chosen, based upon  $P$  (that is, upon all preceding results), to minimize the expected phase variance after the next detection<sup>18</sup>.

This Bayesian control algorithm reduces to Kitaev's algorithm for  $M = 1$ . We have shown analytically that this algorithm gives a standard deviation of estimates scaling as the SQL for  $M = 2$  as well as  $M = 1$ , but numerical simulations (for  $N$  up to  $4 \times 10^6$  and  $M$  up to 8) demonstrate a Heisenberg-limited scaling for  $M \geq 4$ .

We note that a single photon with  $p$  passes through the unknown phase shift is operationally equivalent to a NOON state with  $n = p$

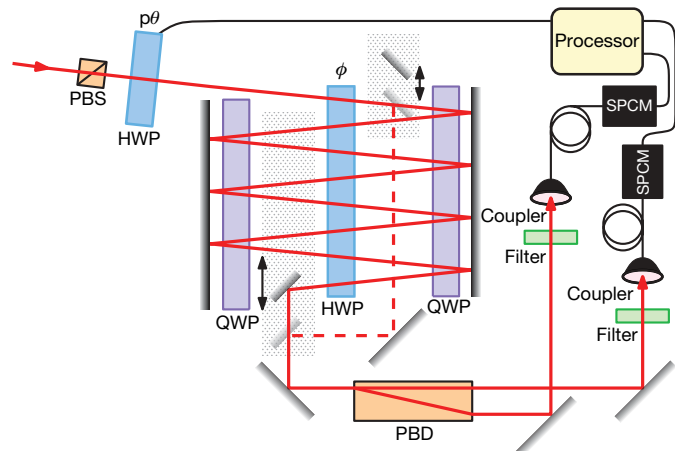


**Figure 2 | Conceptual diagram of the algorithm's implementation as a Mach-Zehnder interferometer.** This is equivalent to the scheme in Fig. 1c. Photon-number quantum states are shown at key points. The first beam splitter implements the Hadamard operation on incident photons. The large phase-shift element is configured to implement a  $p\phi$  phase shift on logical  $|1\rangle$  states, with  $p$  adjustable ( $p = 8$  shown). The small phase-shift element implements the adjustable  $p\theta$  phase shift on logical  $|0\rangle$  states. The final beam splitter and single photon detectors implement a  $\sigma_x$  measurement, which determines, via the processor, how to adjust  $\theta$  before the next photon input.

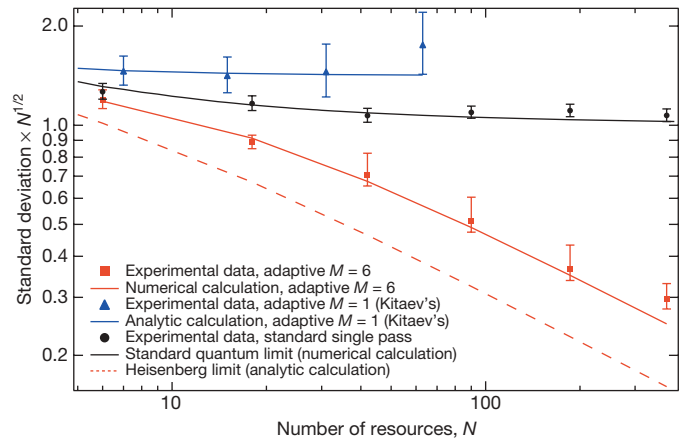
photons, and involves exactly the same number of resources. A single NOON state such as this yields at most one bit of information<sup>4</sup>, and only about  $\phi$  modulo  $2\pi/n$ . It has been shown numerically<sup>21</sup> that a sequence of NOON states, with  $n$  as well as  $\theta$  chosen adaptively, achieves Heisenberg-limited scaling, but only for  $N > 100$ . Our generalized algorithm, which is simpler, can also be applied to NOON states, and directly achieves Heisenberg-limited scaling. Even if high- $n$  NOON states could be produced, however, they require high- $n$  photon-number-resolving detectors, and are proportionately more sensitive to detector inefficiency than single photons with multiple passes through the phase shift.

We demonstrate our single-photon algorithm using a common-spatial-mode polarization interferometer, as shown in Fig. 3. Common-spatial-mode interferometers are used in many metrology tasks involving birefringent materials such as stress sensors, Faraday spectroscopy, and testing optical components, but we stress that our algorithm applies equally well to any interferometer. The two arms of the interferometer are the right-circular ( $|R\rangle$ ) and left-circular ( $|L\rangle$ ) polarization modes. The unknown phase  $\phi$  is implemented as a birefringent half-wave plate. We test two versions of the algorithm:  $M = 1$  (Kitaev's algorithm), and  $M = 6$  (chosen for its robustness). For each, we vary the number of resources  $N = M(2^{K+1} - 1)$  by choosing different values for the maximum number of passes,  $2^K$ , with  $K \in \{0, 1, 2, 3, 4, 5\}$ . We also measured the standard deviation for a non-adaptive or 'standard' estimation algorithm, using  $N$  single passes of the phase shift, with  $N$  chosen equal to the number of resources used for each of the  $M = 6$  data points. In this case  $\theta$  was incremented non-adaptively<sup>29</sup> by  $\pi/N$  from one photon to the next, to ensure a sensitivity independent of  $\phi - \theta_{\text{init}}$ .

The experimental results are shown in Fig. 4, together with theoretical calculations. The error bars are 95% confidence intervals determined using a studentized bootstrap on a log scale<sup>30</sup>. In general, the distributions have a large positive kurtosis which emphasizes the effect of outliers; our error calculation takes this into account to provide accurate error bars. Theoretical predictions assume 100% visibility. Experimentally, visibilities for  $p = 1$  to 16 were high (all above 98.1% and typically above 99.6%). However the  $p = 32$  case had slightly lower visibility (95.4%), leading to higher than expected standard deviations for the  $N = 378$  case. This is primarily due to



**Figure 3 | Schematic of the experiment.** Polarization modes replace the arms of the interferometer in Fig. 2, with phase shifts implemented by half-wave plates (HWPs). A photon experiences phase shifts between left- and right-circular polarizations by the feedback wave plate ( $p\theta$ ) and the unknown phase wave plate ( $\phi$ ). The photon is selected by a mirror mounted on a motorized translation stage, discriminated in the horizontal/vertical polarization basis by a polarizing beam displacer (PBD), and passed through a 10-nm-bandwidth interference filter. It is then coupled into a multimode fibre and detected by a single-photon counting module (SPCM), completing the  $\sigma_x$  measurement. PBS, polarizing beam splitter; QWP, quarter-wave plate.



**Figure 4 | Standard deviations of distributions of phase estimates for varying numbers of resources  $N$ .** We compare theoretical predictions (lines) and measured values (points, each representing 1,000 estimates) for standard phase estimation and our implementation of Kitaev's ( $M = 1$ ) and generalized Kitaev's ( $M = 6$ ) algorithms. Error bars denote 95% confidence intervals. Our algorithm clearly has a lower standard deviation in phase estimates than both the SQL and Kitaev's algorithm (which has SQL scaling). For large  $N$ , the curve for the adaptive algorithm is parallel to the Heisenberg limit, with a small overhead factor of about 1.56.

expansion of the beam, with consequent overlap of beams, leading to a small probability of measuring the photon after only 30 passes.

The results of the non-adaptive phase estimation algorithm follow the SQL, as expected. We note that the standard deviations of the  $M = 1$  (Kitaev's) case also follow an SQL scaling. Most importantly, there is a clear Heisenberg scaling,  $\Delta\phi_{\text{est}} \propto 1/N$ , of our adaptive multi-pass algorithm for  $M = 6$ . Our data are consistent with the predicted overhead factor of 1.56 relative to the asymptotic Heisenberg-limited standard deviation  $\pi/N$ . Despite this constant overhead, our phase estimates clearly surpass the SQL. For example, where we have demonstrated the use of 378 resources in our algorithm (corresponding to a maximum of 32 passes), 4,333 resources would be required using standard techniques to achieve the same uncertainty.

We have introduced a new algorithm for phase estimation, generalizing Kitaev's algorithm, which requires no entanglement to achieve Heisenberg-limited scaling independent of  $\phi$ . Our algorithm uses single-photon Fock states, multiple passes and adaptive measurement. We have used our algorithm successfully to demonstrate the first measurement with Heisenberg-limited scaling. This technique has promise for a wide range of metrology tasks, especially in the light of continued development of high-flux single photon sources and efficient detectors.

## METHODS SUMMARY

**Interferometer.** A spontaneous parametric downconverter supplies pairs of single photons, one to the interferometer and one to an independent detector. The state after the polarizing beam splitter is  $(|R\rangle + |L\rangle)/\sqrt{2}$ , equivalent to the logical state  $H|0\rangle$ . A photon passing through the 50-mm-diameter  $\phi$  phase shift half-wave plate undergoes a polarization rotation:  $\frac{1}{\sqrt{2}}(|R\rangle + |L\rangle) \rightarrow \frac{1}{\sqrt{2}}(|R\rangle + e^{i\phi}|L\rangle)$  for a half-wave plate setting of  $\phi/4$ . Two 50-mm-diameter mirrors are placed on either side, allowing a single photon to pass through the half-wave plate multiple times. To correct the unwanted  $\pi$ -phase shift (in the  $|H\rangle|V\rangle$  basis) on reflection, a quarter-wave plate, set to its optic axis, is inserted before each of the large mirrors. The feedback phase is implemented as another half-wave plate mounted in a computer-controlled rotation stage before the unknown phase and mirrors. We use a fixed phase  $\phi$ , but the use of a uniformly distributed random initial feedback phase is equivalent to performing the protocol over the full range of system phases,  $\phi \in [0, 2\pi)$ . Mirrors on computer-controlled translation stages are used to select the  $2^k$ th pass for each value of  $k$ . Measurement is performed in the horizontal/vertical basis, corresponding to a  $\sigma_x$  measurement, with a high-contrast-ratio calcite polarizing beam displacer. The two outputs of the beam displacer, filtered with 10-nm-bandwidth



filters to reject background light, are sent to single photon counting modules. A successful measurement is heralded by a coincidence between the directly detected photon and either output detector.

**Statistics for phase.** For a phase  $\phi$  with distribution  $P$ , an appropriate measure of error in the estimate  $\phi_{\text{est}}$  is  $\langle \cos(\phi - \phi_{\text{est}}) \rangle_P^{-2} - 1$ . This achieves its minimum, the Holevo variance  $V_H$  (ref. 28), for  $\phi_{\text{est}} = \arg(\langle \exp(i\phi) \rangle_P)$ , which is the estimate we use. When applied to a phase, the terms variance and standard deviation are to be understood as  $V_H$  and  $\sqrt{V_H}$  respectively.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 1 August; accepted 11 September 2007.**

1. Caves, C. M. Quantum-mechanical noise in an interferometer. *Phys. Rev. D* **23**, 1693–1708 (1981).
2. Yurke, B., McCall, S. L. & Klauder, J. R. SU(2) and SU(1,1) interferometers. *Phys. Rev. A* **33**, 4033–4054 (1986).
3. Giovannetti, V., Lloyd, S. & Maccone, L. Quantum-enhanced measurements: beating the standard quantum limit. *Science* **306**, 1330–1336 (2004).
4. Bollinger, J. J., Itano, W. M., Wineland, D. J. & Heinzen, D. J. Optimal frequency measurements with maximally correlated states. *Phys. Rev. A* **54**, R4649–R4652 (1996).
5. Lee, H., Kok, P. & Dowling, J. P. A quantum Rosetta stone for interferometry. *J. Mod. Opt.* **49**, 2325–2338 (2002).
6. Rarity, J. G. *et al.* Two-photon interference in a Mach-Zehnder interferometer. *Phys. Rev. Lett.* **65**, 1348–1351 (1990).
7. Fonseca, E. J. S., Monken, C. H. & Pádua, S. Measurement of the de Broglie wavelength of a multiphoton wave packet. *Phys. Rev. Lett.* **82**, 2868–2871 (1999).
8. Edamatsu, K., Shimizu, R. & Itoh, T. Measurement of the photonic de Broglie wavelength of entangled photon pairs generated by spontaneous parametric down-conversion. *Phys. Rev. Lett.* **89**, 213601 (2002).
9. Walther, P. *et al.* De Broglie wavelength of a non-local four-photon state. *Nature* **429**, 158–161 (2004).
10. Mitchell, M. W., Lundeen, J. S. & Steinberg, A. M. Super-resolving phase measurements with a multiphoton entangled state. *Nature* **429**, 161–164 (2004).
11. Eisenberg, H. S., Hodelin, J. F., Khoury, G. & Bouwmeester, D. Multiphoton path entanglement by nonlocal bunching. *Phys. Rev. Lett.* **94**, 090502 (2005).
12. Leibfried, D. *et al.* Creation of a six-atom 'Schrödinger cat' state. *Nature* **438**, 639–642 (2005).
13. Sun, F. W., Liu, B. H., Huang, Y. F., Ou, Z. Y. & Guo, G. C. Observation of the four-photon de Broglie wavelength by state-projection measurement. *Phys. Rev. A* **74**, 033812 (2006).
14. Nagata, T., Okamoto, R., O'Brien, J. L., Sasaki, K. & Takeuchi, S. Beating the standard quantum limit with four-entangled photons. *Science* **316**, 726–729 (2007).
15. Resch, K. J. *et al.* Time-reversal and super-resolving phase measurements. *Phys. Rev. Lett.* **98**, 223601 (2007).
16. Kitaev, A. Y. Quantum measurements and the Abelian stabilizer problem. *Electr. Coll. Comput. Complex.* **3**, article no. 3 (1996).
17. Wiseman, H. M. Adaptive phase measurements of optical modes: Going beyond the marginal Q distribution. *Phys. Rev. Lett.* **75**, 4587–4590 (1995).
18. Berry, D. W. & Wiseman, H. M. Optimal states and almost optimal adaptive measurements for quantum interferometry. *Phys. Rev. Lett.* **85**, 5098–5101 (2000).
19. Armen, M. A., Au, J. K., Stockton, J. K., Doherty, A. C. & Mabuchi, H. Adaptive homodyne measurement of optical phase. *Phys. Rev. Lett.* **89**, 133602 (2002).
20. Cook, R. L., Martin, P. J. & Geremia, J. M. Optical coherent state discrimination using a closed-loop quantum measurement. *Nature* **446**, 774–777 (2007).
21. Mitchell, M. W. Metrology with entangled states. *Proc. SPIE* **5893**, 589310 (2005).
22. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, Cambridge, UK, 2000).
23. Rudolph, T. & Grover, L. Quantum communication complexity of establishing a shared reference frame. *Phys. Rev. Lett.* **91**, 217905 (2003).
24. de Burgh, M. & Bartlett, S. D. Quantum methods for clock synchronization: beating the standard quantum limit without entanglement. *Phys. Rev. A* **72**, 042301 (2005).
25. Giovannetti, V., Lloyd, S. & Maccone, L. Quantum metrology. *Phys. Rev. Lett.* **96**, 010401 (2006).
26. van Dam, W., D'Ariano, G. M., Ekert, A., Macchiavello, C. & Mosca, M. Optimal quantum circuits for general phase estimation. *Phys. Rev. Lett.* **98**, 090501 (2007).
27. Griffiths, R. B. & Niu, C.-S. Semiclassical Fourier transform for quantum computation. *Phys. Rev. Lett.* **76**, 3228–3231 (1996).
28. Wiseman, H. M. & Killip, R. B. Adaptive single-shot phase measurements: a semiclassical approach. *Phys. Rev. A* **56**, 944–957 (1997).
29. Hradil, Z. *et al.* Quantum phase in interferometry. *Phys. Rev. Lett.* **76**, 4295–4298 (1996).
30. Davison, A. C. & Hinkley, D. V. *Bootstrap Methods and Their Application* Ch. 5 Cambridge Univ. Press, Cambridge, UK (1997).

**Acknowledgements** We thank M. Mitchell, D. Bulger and S. Lo for discussions. This work was supported by the Australian Research Council and the Queensland State Government.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to G.J.P. ([g.pryde@griffith.edu.au](mailto:g.pryde@griffith.edu.au)).

## METHODS

**Source.** Our type-I BiBO (bismuth borate) spontaneous parametric down-conversion source is pumped by a frequency-doubled mode-locked Ti:sapphire laser, producing pairs of 820-nm, 2-nm-bandwidth single photons in the state  $|HH\rangle$ . One photon is guided to the experiment through a single-mode optical fibre, the other is guided straight to a single-photon counting module. A successful measurement is heralded by a coincidence between the directly detected photon and either of the output detectors—coincidence detection reduces background and dark counts, ensuring high-fidelity conditional single photons in the experiment.

**Quarter-wave plate setting.** For logistical reasons, we use 25-mm-diameter quarter-wave plates for most experiments, but 50-mm-diameter wave plates for the  $K=5$  cases. The wave plates are nominally identical except for their diameter. The quarter-wave plates did introduce a technical challenge: because of the large-diameter optics, and the need to use all of the clear aperture to obtain multiple reflections, we used mountings that did not allow easy calibration and adjustment of the quarter-wave plates. This in turn led to small additional phase shifts from the quarter-wave plates that were dependent upon the number of passes. This problem is easily modelled and is not fundamental.

**Analytic solutions.** For the cases  $M=1$  (Kitaev) and  $M=2$  of our algorithm we have shown analytically that the variance scales as the SQL, by solving the adaptive scheme exactly, using the formulae in ref. 18. The exact results for the variances are  $2/N + 1/N^2$  and  $2/N$  respectively.

**Error calculation.** The 95% confidence intervals shown were determined using a studentized bootstrap on a log scale. The bootstrap is a method of determining confidence intervals without making assumptions about the form of the underlying distribution<sup>30</sup>. The data are used as a model of the underlying distribution, and confidence intervals for the quantity of interest are estimated by sampling from this distribution. That is, a number of subsamples equal to the size of the data set,  $m$ , is obtained, and an estimate of the quantity of interest is determined from this set of subsamples. A large number of bootstrap samples are used (in our case  $10^6 - 1$ ), where each sample is an estimate of the quantity of interest based on the set of  $m$  subsamples. The distribution obtained for the quantity of interest is then used to determine a confidence interval, as described on page 199 of ref. 30.

For accuracy, the quantity of interest should have an uncertainty which is independent of the value of the quantity. As the uncertainty in an estimate of variance is approximately proportional to the variance, taking the logarithm of the variance yields a quantity with constant uncertainty (as is done, for example, in ref. 31). The studentized bootstrap yields additional accuracy, and involves normalizing by the estimated uncertainty in the quantity of interest<sup>30</sup>. We have used both the log scale and the studentized bootstrap to obtain accurate error bars.

31. Schenker, N. Qualms about bootstrap confidence intervals. *J. Am. Stat. Assoc.* **80**, 360–361 (1985).

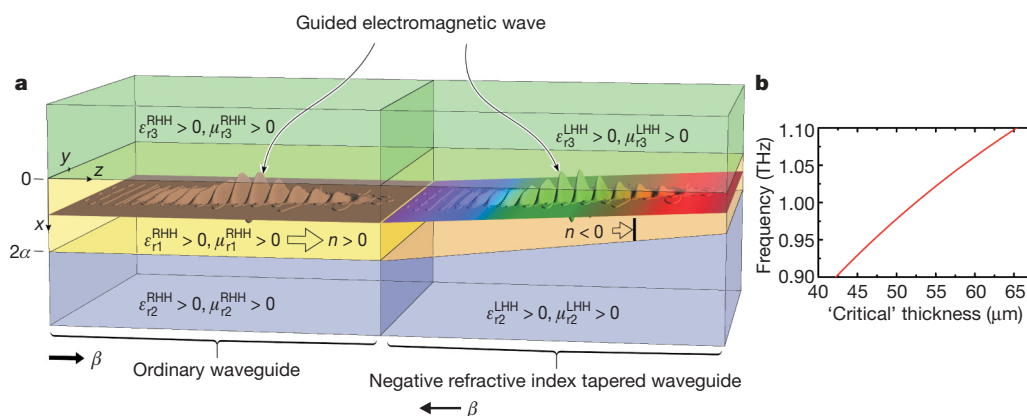
# 'Trapped rainbow' storage of light in metamaterials

Kosmas L. Tsakmakidis<sup>1</sup>, Allan D. Boardman<sup>2</sup> & Ortwin Hess<sup>1</sup>

Light usually propagates inside transparent materials in well known ways<sup>1</sup>. However, recent research<sup>2–6</sup> has examined the possibility of modifying the way the light travels by taking a normal transparent dielectric and inserting tiny metallic inclusions of various shapes and arrangements. As light passes through these structures, oscillating electric currents are set up that generate electromagnetic field moments; these can lead to dramatic effects on the light propagation, such as negative refraction. Possible applications include lenses that break traditional diffraction limits<sup>3,4</sup> and 'invisibility cloaks' (refs 5, 6). Significantly less research has focused on the potential of such structures for slowing, trapping and releasing light signals. Here we demonstrate theoretically that an axially varying heterostructure with a metamaterial core of negative refractive index can be used to efficiently and coherently bring light to a complete standstill. In contrast to previous approaches for decelerating and storing light<sup>7–13</sup>, the present scheme simultaneously allows for high in-coupling efficiencies and broadband, room-temperature operation. Surprisingly, our analysis reveals a critical point at which the effective thickness of the waveguide is reduced to zero, preventing the light wave from propagating further. At this point, the light ray is permanently

trapped, its trajectory forming a double light-cone that we call an 'optical clepsydra'. Each frequency component of the wave packet is stopped at a different guide thickness, leading to the spatial separation of its spectrum and the formation of a 'trapped rainbow'. Our results bridge the gap between two important contemporary realms of science—metamaterials and slow light. Combined investigations may lead to applications in optical data processing and storage or the realization of quantum optical memories.

For decades scientists maintained that<sup>14</sup> "optical data cannot be stored statically and must be processed and switched on the fly". The reason for this conclusion was that stopping and storing an optical signal by dramatically reducing the speed of light itself was thought to be infeasible. Undeniably, the absence of any form of interaction between photons and other elementary particles, as well as their enormous speed, makes it excessively difficult to confine them to a finite volume by reducing their velocity down to zero. However, in recent years a series of discoveries contributed towards annulling such assertions and proved conclusively that it is, indeed, possible to bring light to a complete standstill. Amongst others, electro-magnetically induced transparency<sup>7</sup>, quantum-dot semiconductor



**Figure 1 | Trapped rainbow.** Different frequency components of a guided wave packet stop at correspondingly different thicknesses inside a tapered left-handed heterostructure (LHH). **a**, The thick open arrows reveal the direction of power flow propagation ( $P_{\text{tot}}^z$ ), while the thin black arrows show the direction of phase propagation ( $\beta$ ). The numbering of the layers follows the magnitude of the corresponding refractive indices—that is, in both heterostructures, the middle layer (1) has the highest magnitude of refractive index, followed by the lower layer (2) and, last, by the upper layer (3). A guided wave packet is efficiently injected from the ordinary waveguide to the LHH (see also Fig. 4), inside which it propagates smoothly owing to the slow (adiabatic) reduction in the thickness of the core. The smallest (red) frequency components of the wave are stopped at the smallest core thicknesses of the LHH, while the largest (blue) components stop at correspondingly larger core thicknesses. Thereby, the spectrum of the oscillatory field will be spatially decomposed into its frequency constituents.

After it is trapped, a light ray can be released by locally tuning in real-time the microphotonic structure<sup>4,9,30</sup> to perturb the repeating 'optical clepsydra' pattern of the ray (Fig 3b–e), and by reversely adiabatically tapering the LHH back to the wide thickness point, which can be impedance-matched to an out-coupling RHH (see Fig. 4). The path of the forward propagating ray inside the reversely tapered LHH (RT-LHH) will be the mirror image of a corresponding backwards propagating ray inside the depicted LHH. From the law of reversibility of the light rays, one then sees that the forward propagating ray inside the RT-LHH, after it has traced its dual path, will reach the wide-thickness end of the structure and will enter the out-coupling RHH with minimal reflection, just as light was initially injected from the in-coupling RHH to our tapered LHH. **b**, An example of the dependence of the 'critical' thicknesses of the LHH on the spectrum of the wave packet. The optogeometric parameters used in this example are those of Fig. 2.

<sup>1</sup>Advanced Technology Institute and Department of Physics, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford GU7 1QR, UK. <sup>2</sup>Photonics and Nonlinear Science Group, Joule Laboratory, Department of Physics, University of Salford, Salford M5 4WT, UK.

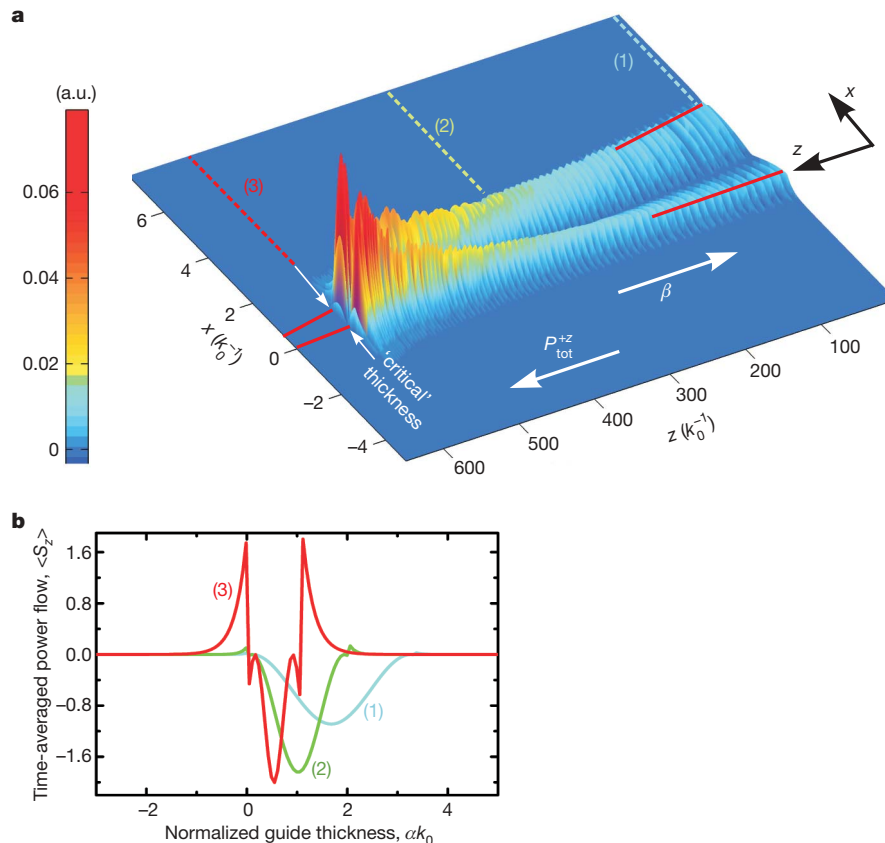


optical amplifiers<sup>8</sup>, photonic crystals<sup>9,15</sup>, coherent population oscillations<sup>10</sup>, stimulated Brillouin scattering<sup>11</sup> and surface plasmon polaritons<sup>12,13</sup> in metalodielectric waveguides have been proposed as means of producing 'slow light'. However, so far most of these methods bear inherent limitations that may hinder their practical deployment. For instance, electromagnetically induced transparency uses ultracold atomic gases and not solid state materials, quantum-dot semiconductor optical amplifiers usually allow for only modest delays (but for potentially ultra-broadband light pulses), coherent population oscillations and stimulated Brillouin scattering are very narrowband owing to the narrow transparency window of the former and the narrow Brillouin gain bandwidth (around 30 MHz in standard single-mode optical fibres) of the latter, and surface plasmon polaritons are very sensitive to surface roughness and are relatively difficult to excite. Photonic crystals are normally highly multimodal<sup>15</sup>; this, combined with the strong impedance mismatch in the 'slow-light regime', makes it overly difficult to launch the incoming light energy to a single, slow mode alone<sup>16</sup>.

During the same period and in parallel with the above advances, a different, wide-ranging realm of contemporary research has also been developing. It largely followed from a sequence of works by Pendry, Smith and co-workers, in which they proposed practical means for realizing negative refractive index metamaterials<sup>17</sup> and meticulously demonstrated their operation<sup>18</sup>. These materials offer a new perspective to the optical world and provide additional degrees of freedom in the design of photonic devices<sup>5</sup>, thereby allowing unprecedented control over the flow of light. A 'perfect' lens<sup>3</sup>, highly anisotropic electromagnetic 'cloaks' that render objects invisible to incident

radiation<sup>6</sup>, as well as focusing of electron de Broglie waves by sharp p-n junctions in graphene<sup>19</sup>, are a few characteristic examples. Moreover, there is now compelling evidence that, by building on familiar transmission-line concepts borrowed from microwave analysis, such materials can be designed to exhibit broadband negative-index behaviour and relative robustness to losses, all through the microwave region and up to the ultraviolet<sup>20,21</sup> domain.

In this work, we bring together the realms of metamaterials and slow light as a result of studying the physics associated with wave propagation in slowly, spatially varying negative-index heterostructures. To gain an insight into the physics of the problem at hand, let us for a moment imagine a ray of light propagating in a zigzag fashion along a waveguide with a negative-index ('left-handed') core. The ray experiences negative Goos-Hänchen lateral displacements<sup>22</sup> each time it strikes the interfaces of the core with the positive-index ('right-handed') claddings. Accordingly, the cross points of the incident and reflected rays will sit inside the left-handed core and the effective thickness of the guide will be smaller than its natural thickness (see below). It is reasonable to expect that by gradually reducing the core physical thickness, the effective thickness of the guide will eventually vanish. Obviously, beyond that point the ray will not be able to propagate further down, and will effectively be trapped inside the left-handed heterostructure (LHH). For non-radiative trapping of the ray, we hope that the point where the effective thickness vanishes will not occur below the lower cut-off frequency of the guide. If this condition is fulfilled, then one of the frequencies of a guided wave packet will be completely arrested. The arresting points for the other frequency components will be continuously spaced out. Thereby, one



**Figure 2 | Wave propagation in an axially varying LHH.** The optical parameters of the structure are  $\epsilon_{r1}^{\text{LHH}} = -5$ ,  $\epsilon_{r2}^{\text{LHH}} = 2.56$ ,  $\epsilon_{r3}^{\text{LHH}} = 2.25$ ,  $\mu_{r1}^{\text{LHH}} = -5$  and  $\mu_{r2}^{\text{LHH}} = \mu_{r3}^{\text{LHH}} = 1$ . **a**, A snapshot of the propagation of the monochromatic ( $f = 1$  THz) p-polarized magnetic-field component, which enters the LHH from the wide-thickness end ( $\alpha k_0 = 1.7$ ) and stops at the pre-arranged 'critical' thickness of  $\alpha k_0 \approx 0.55$ . The light signal carries a total power  $P_{\text{tot}}^{\text{con}} = 82.6 \mu\text{W m}^{-2} \text{s}^{-1}$ , which to an excellent approximation is

assumed to be unaffected by the negligibly small reflections from the approaching media interfaces. The group velocity of the wave is parallel to the total power flow  $P_{\text{tot}}^{+z}$ , while the phase velocity is parallel to the longitudinal propagation constant  $\beta$ . **b**, The evolution of the (discontinuous) time-averaged power flow<sup>24,26</sup>  $\langle S_z \rangle = [\beta / (2\omega\epsilon_0\epsilon_i)] \times |H_y|^2$  from the wide-thickness end (blue line) to the 'critical' thickness (red line).

will in principle have a means of stopping light in its tracks and storing it indefinitely, without radiation, over a whole range of frequencies. Such an approach will also be totally immune to atomic decoherence effects<sup>23</sup> and will not require atomic media or group index resonances.

In the following, we show that efficient, broadband, non-radiative storage of light can be realized inside an LHH. We consider an axially non-uniform, linearly tapered, planar waveguide with a core of isotropic and homogeneous negative-index metamaterial, bounded asymmetrically by two positive-index media, as illustrated in Fig. 1a. The simultaneous negativity of the permittivity  $\epsilon$  and permeability  $\mu$  in the left-handed core allows for the existence of efficiently excitable (oscillatory) waveguide modes<sup>24</sup>, which are not supported by single-negative plasmonic structures. We arrange the variation of the core half-thickness  $\alpha$  with distance  $z$  to be appropriately slow<sup>25</sup> (adiabatic),  $d\alpha/dz < 0.05 \times \min\{(\alpha k_0) \times (n_{\text{eff},2} - n_{\text{eff},3})/(2\pi)\}$ , with  $\alpha k_0$  being the reduced guide thickness,  $k_0$  the free space wavenumber and  $n_{\text{eff},2}$ ,  $n_{\text{eff},3}$  the effective indices of the second- and third-order oscillatory modes<sup>24</sup>, so that the power of a local mode is conserved along the structure. One may then analytically show (see Supplementary Information) that the

electromagnetic fields at a position  $z_t$  are given by:

$$\mathbf{G}(x, y, z_t, t) = F(z_t) \mathbf{g}(x, y, \beta(z_t)) \exp \left[ i \sum_{z=0}^{z=z_t} \beta(z) \Delta z - i \omega t \right] \quad (1)$$

Here,  $\mathbf{G} = \mathbf{E}$  or  $\mathbf{H}$ ,  $\mathbf{g}$  is the solution to the  $E$ - or  $H$ -field vector wave equation for  $z = z_p$ ,  $\beta$  is the longitudinal propagation constant,  $\omega$  is the wave angular frequency and  $F$  is an appropriate factor, which carries all the information about the mode energy conservation and is calculated as:

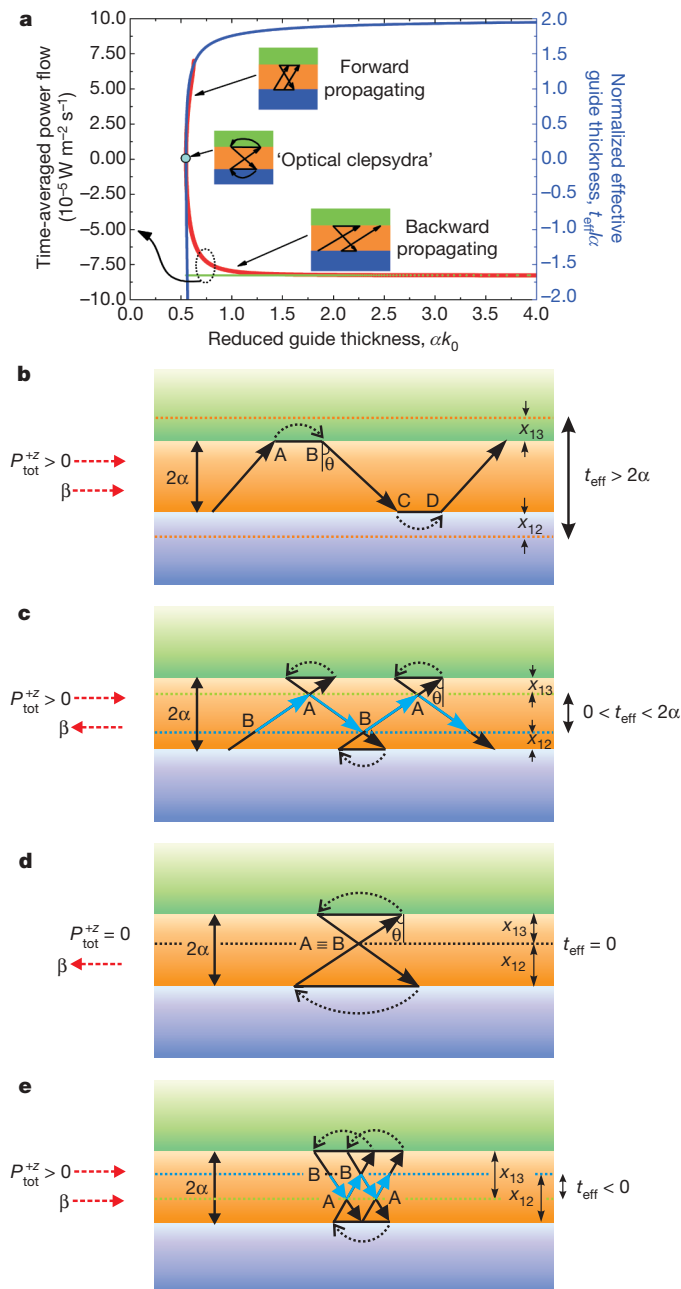
$$F = 2\sigma_e U \left[ \frac{\omega \epsilon_0 |\epsilon_{r1}| P_{\text{tot}}^{\text{con}}}{\alpha \beta (W_3^2 + \sigma_e^2 U^2) \Theta} \right]^{1/2} \quad (2)$$

where  $\sigma_e = \epsilon_{r3}/|\epsilon_{r1}|$ ,  $\epsilon_{r3}$  and  $\epsilon_{r1}$  are the relative permittivities of the upper (3) and middle (1) layers, respectively,  $U = \alpha \kappa$  and  $W_3 = \alpha \gamma_3$  are the reduced transverse and upper-layer decay constants, respectively ( $\kappa$  and  $\gamma_3$  are the corresponding unnormalized constants),  $P_{\text{tot}}^{\text{con}} = |P_1| + |P_2| + |P_3|$  is the conserved total time-averaged power flow, with  $P_i$  ( $i = 1, 2, 3$ ) being the time-averaged power flow in layer  $i$ , propagating in the increasing  $z$  direction, and:

$$\Theta = 2 + \frac{\rho_e}{W_2} \frac{U^2 - W_2^2}{W_2^2 + \rho_e^2 U^2} + \frac{\sigma_e}{W_3} \frac{U^2 - W_3^2}{W_3^2 + \sigma_e^2 U^2} \quad (3)$$

with  $\rho_e = \epsilon_{r2}/|\epsilon_{r1}|$  and  $W_2 = \alpha \gamma_2$  being the reduced decay constant in the lower layer ( $\gamma_2$  is the corresponding unnormalized decay constant for the lower layer).

One can judiciously choose the optical parameters of the LHH to facilitate total suppression of all surface polariton modes—that is, the minimum thickness of the core layer can be chosen to be above the upper cut-off of the parasitic surface polariton mode, so that only oscillatory guided modes may exist<sup>26</sup>. For such a structure, Fig. 1b illustrates an example of ‘critical’ guide thicknesses within which a wave packet, having a spectrum centred at 1 THz, can be completely stopped. Further, Fig. 2 furnishes *ab initio* calculations, following the previously outlined methodology, of the propagating, monochromatic, p-polarized magnetic-field component at a time instant  $t = 0$ . We find that the



**Figure 3 | Ray analysis reveals that the effective left-handed guide thickness is smaller than the physical thickness and can become zero or even negative.** **a**, Variations of normalized effective guide thickness  $t_{\text{eff}}/\alpha$  (solid blue line), conserved  $-P_{\text{tot}}^{\text{con}}$  (dotted green line) and forward  $P_{\text{tot}}^{+z}$  (solid red line) total time-averaged power flow with the reduced guide thickness  $\alpha k_0$ . The forward component (red) of the conserved power flow (green) gradually decreases in magnitude, until it becomes exactly zero at the critical core thickness. At this point, the effective thickness of the LHH also vanishes. For larger guide thicknesses,  $t_{\text{eff}}/\alpha$  tends asymptotically to the value of 2. The insets associate characteristic regions of  $P_{\text{tot}}^{+z}$  with the ray analysis results shown in **c–e**. In **b–e**, a ray of light (black), which here signifies power propagation, hits the media interfaces with an angle  $\theta$  while propagating down the waveguide, and experiences a Goos-Hänchen lateral displacement. The black dotted arrows denote the evanescent field power flow from the optically denser core to the rarer claddings<sup>25</sup>. In a normal dielectric waveguide<sup>25,28</sup> (**b**), the Goos-Hänchen phase shifts are positive, the core appears to extend (dotted orange lines) inside the cladding layers and the effective guide thickness is always larger than the physical core thickness  $2\alpha$ . In the slowly varying LHH case (**c–e**), the thickness of the core remains practically constant over many ray periods owing to the slow variation criterion. Here, the ray experiences a negative, that is antiparallel to  $P_{\text{tot}}^{+z}$ , later displacement originating from the reversed power flow from the core to the claddings and the associated negative Goos-Hänchen phase shift. The two cladding layers now appear to extend inside the core layer (dotted blue and dotted green lines in **c, e**). In **c**, the shifts are relatively small, such that  $x_{12} + x_{13} < 2\alpha$ . Accordingly, the ray (blue) is effectively confined in the middle region of thickness  $t_{\text{eff}}$ , by repeatedly bouncing off points A and B. For an appropriate choice of optogeometric parameters (**d**), the two phase shifts can become such that  $x_{12} + x_{13} = 2\alpha$  exactly. In this case,  $t_{\text{eff}}$  vanishes and the ray becomes permanently trapped, forming a double light cone (‘optical clepsydra’). For even larger, compared to the ray period, later displacements (**e**),  $t_{\text{eff}}$  becomes negative; the ray is still guided in the middle region of thickness  $|t_{\text{eff}}|$ , but now is forward-propagating, that is, the direction of phase propagation is parallel to the direction of power flow.

magnitude of the total time-averaged power flow propagating in the positive  $+z$  direction,  $P_{\text{tot}}^{+z} = 1/2 \int_{-\infty}^{\infty} \text{Re}(\mathbf{E} \times \mathbf{H}^*)_z dx$ , gradually drops off until it totally peters out (Figs 2b and 3a), even though  $P_{\text{tot}}^{\text{con}}$  is conserved along the non-uniform waveguide. Accordingly, one discovers that while the guided oscillatory fields propagate along the structure, with their phase ( $v_{\text{ph}}$ ) and group ( $v_g$ ) velocities being antiparallel, the group and energy ( $v_E$ ) velocities<sup>27</sup> progressively decrease, eventually becoming zero at a ‘critical’, pre-determined, guide thickness. At this point, the fields are slowly spatially compressed and amassed<sup>23</sup>, with their total amplitude increasing by a factor of 4 for the present set of optogeometric parameters. The exponentially decaying extension of the field inside the lower positive-index layer (medium 3) also progressively increases, as anticipated. Note that in a normal dielectric waveguide the monochromatic oscillatory field would have been reflected and radiated off at the guide cut-off thickness. In that case, an increase in the total amplitude of the field would have been the result of interference between the two light waves propagating in opposite directions. In the present LHH case, however, the wave travels solely in the positive  $z$  direction and completely stops upon reaching the ‘critical’ guide thickness; hence, neither reflection nor radiation or interference occur.

The trapping of light is more intuitively recognized by tracing the trajectory of a light ray inside the core for guide thicknesses nearly equal to the critical one mentioned above (see Fig. 3). Recall that the variation of the core refractive index with distance  $z$  is much smaller than the ray half-period  $z_{\text{hp}}$ , so that locally the guide appears practically uniform. Let us assume that the light ray arrives at the 1–3 media interface of the LHH with an angle  $\theta$  (Fig. 3c–e). Following reflection from this interface, the ray experiences a Goos-Hänchen phase shift<sup>22</sup>  $\delta_{p13} = -2 \tan^{-1} [W_3/(\sigma_e U)]$ , which is equal in magnitude but opposite in sign to the shift of the corresponding positive-index case. Detailed calculations then reveal that the distance  $x_{p13}$  between the cross point A of the two rays (Fig. 3c–e) and the 1–3 media interface is (see Supplementary Information)  $x_{p13} = (\alpha \sigma_e V_3^2)/[W_3(W_3^2 + \sigma_e^2 U^2)]$ , where  $V_3^2 = U^2 + W_3^2$ . Likewise, the distance  $x_{p12}$  between point B and the 1–2 interface is  $x_{p12} = (\alpha \rho_e V_2^2)/[W_2(W_2^2 + \rho_e^2 U^2)]$ , where  $V_2^2 = U^2 + W_2^2$ . One is, thus, led to discern (Fig. 3c, blue line) that the light ray is effectively altogether confined within the middle region of thickness  $t_{\text{eff}}$  wherein it repeatedly bounces off points A and B.

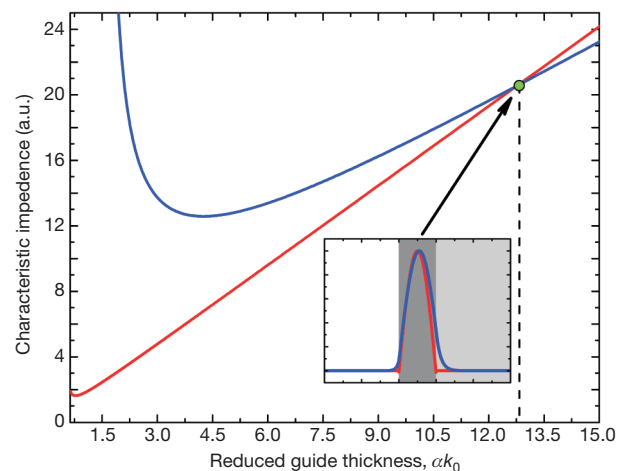
We argue that  $t_{\text{eff}} = 2\alpha - x_{p12} - x_{p13}$  is the effective thickness of the left-handed waveguiding heterostructure. This conclusion is strongly supported by the preceding remarks concerning the ray trajectory, and can be formally established by noting that the total time-averaged power flow  $P_{\text{tot}}^{+z}$  can be directly linked to  $t_{\text{eff}}$  through the following relation (see Supplementary Information):

$$P_{\text{tot}}^{+z} = \frac{1}{4} E_x^{\text{max}} H_y^{\text{max}} t_{\text{eff}} \quad (4)$$

where  $E_x^{\text{max}}$  and  $H_y^{\text{max}}$  are the maximum values of the  $E_x$ - and  $H_y$ -field components in the guide, respectively. The similarities between equation (4) and its counterpart in the case of conventional, right-handed heterostructures (RHH)<sup>28</sup>, allow one to conclusively infer that  $t_{\text{eff}}$ , as defined in equation (4) is, indeed, the effective thickness of the LHH. However, in stark contrast to conventional waveguides,  $t_{\text{eff}}$  is here always smaller than the physical thickness of the core, and can become zero (Fig. 3d) or even negative (Fig. 3e). When  $t_{\text{eff}}$  becomes negative, we deduce from equation (4) that  $P_{\text{tot}}^{+z}$  and  $\beta$  become parallel; thereby, the corresponding guided mode will be a forward-propagating one. Interestingly, on the basis of Fig. 3a and equation (4) we infer that for a particular value of the guide’s physical thickness  $2\alpha$ , the effective thickness vanishes. In this case, the lateral shifts  $x_{p12}$ ,  $x_{p13}$  experienced by the ray upon reflection from the two interfaces are such that  $x_{p12} + x_{p13} = 2\alpha$  exactly. For the aforementioned ‘critical’ physical thickness, the light ray is permanently trapped inside the LHH, being unable to propagate further down. From Fig. 3d we see that in this case the trajectory of the ray forms a double light-cone. In view of its characteristic hourglass form, we will call it the ‘optical clepsydra’.

Following a similar course of analysis, we discover that for different excitation frequencies the guided oscillatory fields stop at correspondingly different guide thicknesses. Accordingly, a guided electromagnetic wave packet can be altogether trapped within a fixed area, spanning a continuous range of guide thicknesses. The leading (trailing) part of the pulse, composed of the smallest (highest) frequencies (Fig. 3a), stops at the smallest (highest) guide thicknesses. Thereby, in the small-intensity, linear case—in which the propagating spectral power densities of a ‘white’ wave packet do not couple<sup>8</sup>, and the guided field is a linear, weighted sum of its single frequency constituents—the ‘red’ and ‘blue’ components of the field will be spatially separated (Fig. 1). This is similar to the separation of the colours of the visible spectrum and the appearance of a rainbow when sunlight illuminates a transparent prism or falling water droplets. For this reason, we shall henceforth call the stopping and storing of light in such LHHs the ‘trapped rainbow’ effect.

A critical characteristic of the axially varying LHH is that further away from the point where the trapping of the light beam is arranged to occur (that is, for larger guide thicknesses) it is possible to achieve complete impedance matching with a dielectric waveguide. We recall that for waveguide structures, the characteristic impedance is defined as  $Z_0^{PV} = (V_0 V_0^*)/(2P_{\text{tot}}^{+z})$ , where  $V_0$  is a ‘voltage’ defined as the line integral of the electric field along some path, which starts from below the lower interface and ends amply above the upper one<sup>29</sup>. Figure 4 illustrates the variation with reduced guide thickness of the analytically calculated (see Supplementary Information) ordinary and left-handed waveguide characteristic impedances. At a point sufficiently far from the ‘critical’ thickness of the LHH, its characteristic impedance becomes equal to that of a normal waveguide. At this point ( $\sim 12.76 \alpha k_0$  in our case) the two structures also have equal thicknesses. Moreover, the spatial distribution of the guided field at the wide end of the LHH (Fig. 4 inset, red line) closely matches the field distribution of a single-mode optical waveguide. Accordingly, a light wave launched from a dielectric guide to a wide-thickness LHH will experience minimal reflection, mainly owing to minute mode-mismatch, which can be further adjusted and optimized at will.



**Figure 4 | Honed conditions for waveguide coupling, showing simultaneous impedance, thickness and mode matching in adjoining RHH and LHHs.** The optical parameters of the LHH are similar to those in Fig. 2. For the RHH we have  $\epsilon_{r1}^{\text{RHH}} = 1.5625$ ,  $\epsilon_{r2}^{\text{RHH}} = 1.44$ ,  $\epsilon_{r3}^{\text{RHH}} = 1.21$ ,  $\mu_{r1}^{\text{RHH}} = \mu_{r2}^{\text{RHH}} = \mu_{r3}^{\text{RHH}} = 1$ . The characteristic impedance of the dielectric waveguide (blue) exhibits a dip at  $\alpha k_0 \approx 4.53$ , after which it grows monotonically. A similar trend is found for the characteristic impedance of the LHH (red). Here, the minimum value occurs at a guide thickness  $\alpha k_0 \approx 0.89$  and at the ‘critical’ LHH thickness it diverges. The two curves cross at  $\alpha k_0 \approx 12.77$ . The inset shows the profile of the fundamental (blue) and second-order oscillatory<sup>24</sup> (red) mode of the dielectric waveguide and LHH, respectively, at the cross point. The darker the shaded region in the inset, the higher is the magnitude of the refractive index.



In summary, we have shown how guided electromagnetic fields can be brought efficiently to a complete standstill while travelling inside axially varying left-handed waveguiding heterostructures. The scheme invokes solid-state materials and, as such, is not subject to low-temperature or atomic coherence limitations. Moreover, it inherently allows for high in-coupling efficiencies and broadband function, as the deceleration of light does not rely on refractive index resonances. This 'trapped rainbow' method for storing photons opens the way to a multitude of hybrid, optoelectronic devices to be used in 'quantum information' processing, communication networks and signal processors, and could lead to combined metamaterials and slow-light research.

Received 5 July; accepted 19 September 2007.

- Born, M. & Wolf, E. *Principles of Optics* 6th edn (Pergamon, Oxford, 1980).
- Grigorenko, A. N. *et al.* Nanofabricated media with negative permeability at visible frequencies. *Nature* **438**, 335–338 (2005).
- Pendry, J. B. Negative refraction makes a perfect lens. *Phys. Rev. Lett.* **85**, 3966–3969 (2000).
- Hou-Tong, C. *et al.* Active terahertz metamaterial devices. *Nature* **444**, 597–600 (2006).
- Pendry, J. B., Schurig, D. & Smith, D. R. Controlling electromagnetic fields. *Science* **312**, 1780–1782 (2006).
- Schurig, D. *et al.* Metamaterial electromagnetic cloak at microwave frequencies. *Science* **314**, 977–980 (2006).
- Liu, C., Dutton, Z., Behroozi, C. H. & Hau, L. V. Observation of coherent optical information storage in an atomic medium using halted light pulses. *Nature* **409**, 490–493 (2001).
- Gehrig, E., van der Poel, M., Mørk, J. & Hess, O. Dynamic spatiotemporal speed control of ultrashort pulses in quantum-dot SOAs. *IEEE J. Quant. Electron.* **42**, 1047–1054 (2006).
- Vlasov, Y. A., O'Boyle, M., Hamann, H. F. & McNab, S. J. Active control of slow light on a chip with photonic crystal waveguides. *Nature* **438**, 65–69 (2005).
- Bigelow, M. S., Lepeshkin, N. N. & Boyd, R. W. Superluminal and slow light propagation in a room-temperature solid. *Science* **301**, 200–202 (2003).
- Okawachi, Y. *et al.* Tunable all-optical delays via Brillouin slow light in an optical fiber. *Phys. Rev. Lett.* **94**, 153902 (2005).
- Stockman, M. I. Nanofocusing of optical energy in tapered plasmonic waveguides. *Phys. Rev. Lett.* **93**, 137404 (2004).
- Karalis, A., Lidorikis, E., Ibanescu, M., Joannopoulos, J. D. & Soljačić, M. Surface-plasmon-assisted guiding of broadband slow and subwavelength light in air. *Phys. Rev. Lett.* **95**, 063901 (2005).
- Blumenthal, D. J., Pruncal, P. R. & Sauer, J. R. Photonic packet switches: Architectures and experimental implementations. *Proc. IEEE* **82**, 1650–1667 (1994).
- Gersen, H. *et al.* Real-space observation of ultraslow light in photonic crystal waveguides. *Phys. Rev. Lett.* **94**, 073903 (2005).
- Vlasov, Y. A. & McNab, S. J. Coupling into the slow light mode in slab-type photonic crystal waveguides. *Opt. Lett.* **31**, 50–52 (2006).
- Pendry, J. B., Holden, A. J., Robbins, D. J. & Stewart, W. J. Magnetism from conductors and enhanced nonlinear phenomena. *IEEE Trans. Microwave Theory Tech.* **47**, 2075–2084 (1999).
- Shelby, R., Smith, D. R. & Schultz, S. Experimental verification of a negative index of refraction. *Science* **292**, 77–79 (2001).
- Cheianov, V. V., Fal'ko, V. & Altshuler, B. L. The focusing of electron flow and a Veselago lens in graphene p-n junctions. *Science* **315**, 1252–1255 (2007).
- Shalaev, V. M. Optical negative-index metamaterials. *Nature Photonics* **1**, 41–48 (2007).
- Alù, A. & Engheta, N. Three-dimensional nanotransmission lines at optical frequencies: A recipe for broadband negative-refraction optical metamaterials. *Phys. Rev. B* **75**, 024304 (2007).
- Berman, P. R. Goos-Hänchen shift in negatively refractive media. *Phys. Rev. E* **66**, 067603 (2002).
- Milonni, P. W. *Fast Light, Slow Light and Left-Handed Light* Chs 5–6 (Institute of Physics, Bristol, 2005).
- Tsakmakidis, K. L., Klaedtke, A., Aryal, D. P., Jamois, C. & Hess, O. Single-mode operation in the slow-light regime using oscillatory waves in generalized left-handed heterostructures. *Appl. Phys. Lett.* **89**, 201103 (2006).
- Snyder, A. & Love, J. D. *Optical Waveguide Theory* Chs 5–19 (Chapman and Hall, New York, 1983).
- Tsakmakidis, K. L., Hermann, C., Klaedtke, A., Jamois, C. & Hess, O. Surface plasmon polaritons in generalized slab heterostructures with negative permittivity and permeability. *Phys. Rev. B* **73**, 085104 (2006).
- Boardman, A. D. & Marinov, K. Electromagnetic energy in a dispersive metamaterial. *Phys. Rev. B* **73**, 165110 (2006).
- Tamir, T. (ed.) *Integrated Optics* Ch. 2 (Springer, New York, 1979).
- Helszajn, J. *Ridge Waveguides and Passive Microwave Components* (IEE Press, London, 2000).
- Povinelli, M. L. *et al.* Evanescent-wave bonding between optical waveguides. *Opt. Lett.* **30**, 3042–3044 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank A. Klaedtke and D. P. Aryal for discussions and technical assistance. This work was supported by the Engineering and Physical Sciences Research Council (UK).

**Author Contributions** K.L.T. and O.H. conceived the presented idea. K.L.T. developed the theory, performed the computations and wrote a draft of the paper. A.D.B. contributed to the discussions. O.H. encouraged K.L.T. to investigate metamaterials and slow light and supervised the findings of the work.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to K.L.T. (K.Tsakmakidis@surrey.ac.uk) or O.H. (O.Hess@surrey.ac.uk).

## LETTERS

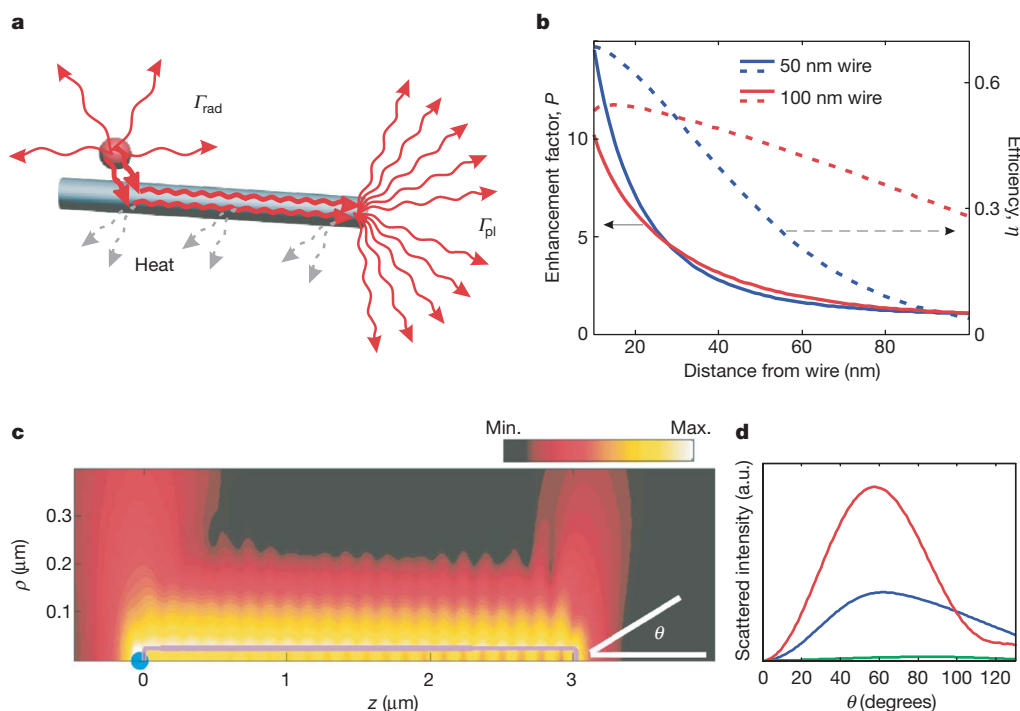
# Generation of single optical plasmons in metallic nanowires coupled to quantum dots

A. V. Akimov<sup>1,4\*</sup>, A. Mukherjee<sup>1\*</sup>, C. L. Yu<sup>2\*</sup>, D. E. Chang<sup>1</sup>, A. S. Zibrov<sup>1,4</sup>, P. R. Hemmer<sup>3</sup>, H. Park<sup>1,2</sup> & M. D. Lukin<sup>1</sup>

Control over the interaction between single photons and individual optical emitters is an outstanding problem in quantum science and engineering. It is of interest for ultimate control over light quanta<sup>1</sup>, as well as for potential applications such as efficient photon collection<sup>2</sup>, single-photon switching<sup>3</sup> and transistors<sup>4</sup>, and long-range optical coupling of quantum bits<sup>5,6</sup>. Recently, substantial advances have been made towards these goals, based on modifying photon fields around an emitter using high-finesse optical cavities<sup>2,3,5–8</sup>. Here we demonstrate a cavity-free, broadband approach for engineering photon–emitter interactions<sup>4,9</sup> via sub-wavelength confinement of optical fields near metallic nanostructures<sup>10–13</sup>. When a single CdSe quantum dot is optically excited in close proximity to a silver nanowire, emission from the quantum dot couples directly to guided surface plasmons in the

nanowire, causing the wire's ends to light up. Non-classical photon correlations between the emission from the quantum dot and the ends of the nanowire demonstrate that the latter stems from the generation of single, quantized plasmons. Results from a large number of devices show that efficient coupling is accompanied by more than 2.5-fold enhancement of the quantum dot spontaneous emission, in good agreement with theoretical predictions.

Surface plasmons, or surface plasmon polaritons, are propagating excitations of charge-density waves and their associated electromagnetic fields on the surface of a conductor<sup>10</sup>. Much like the optical modes of a conventional dielectric fibre, a broad continuum of surface plasmon modes can be confined on a cylindrical metallic wire and guided along the wire axis<sup>12,13</sup> (Fig. 1a). However, as opposed to dielectric waveguides<sup>14,31</sup>, the thin wires can maintain propagation of



**Figure 1 | Radiative coupling of quantum dots to conducting nanowires.** **a**, A coupled quantum dot can either spontaneously emit into free space or into the guided surface plasmons of the nanowire with respective rates  $\Gamma_{\text{rad}}$ ,  $\Gamma_{\text{pl}}$ . **b**, Theoretical dependence of the enhancement factor  $P$  (solid line) and efficiency of emission into surface plasmons (dashed line) on distance of the emitter from the nanowire edge. The red (blue) curve corresponds to a wire diameter of 100 nm (50 nm). **c**, Simulations of the electric field amplitude (arbitrary units) emitted by a dipole (blue filled circle) positioned 25 nm from

one end of a conducting nanowire (whose surface is outlined) 3  $\mu\text{m}$  in length and 50 nm in diameter. The vertical scale ( $\rho$ ) is enlarged compared to the horizontal ( $z$ ) to clearly show the near field of the surface plasmons. Upon hitting the far end of the nanowire, some of the surface plasmon energy is clearly scattered into the far-field, while the remaining is either lost to dissipation or to back-reflection.  $\theta$ , Emission angle. **d**, Amplitude of the Poynting vector of the light scattered from the far end of the nanowire, as a function of  $\theta$  (see **c**), for wires of diameter 100 nm (red curve), 50 nm (blue) and 25 nm (green).

<sup>1</sup>Department of Physics, <sup>2</sup>Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA. <sup>3</sup>Department of Electrical and Computer Engineering, Texas A&M University, College Station, Texas 77843, USA. <sup>4</sup>P.N. Lebedev Physical Institute RAS, Leninskiy prospect 53, Moscow, 119991, Russia.

\*These authors contributed equally to this work.

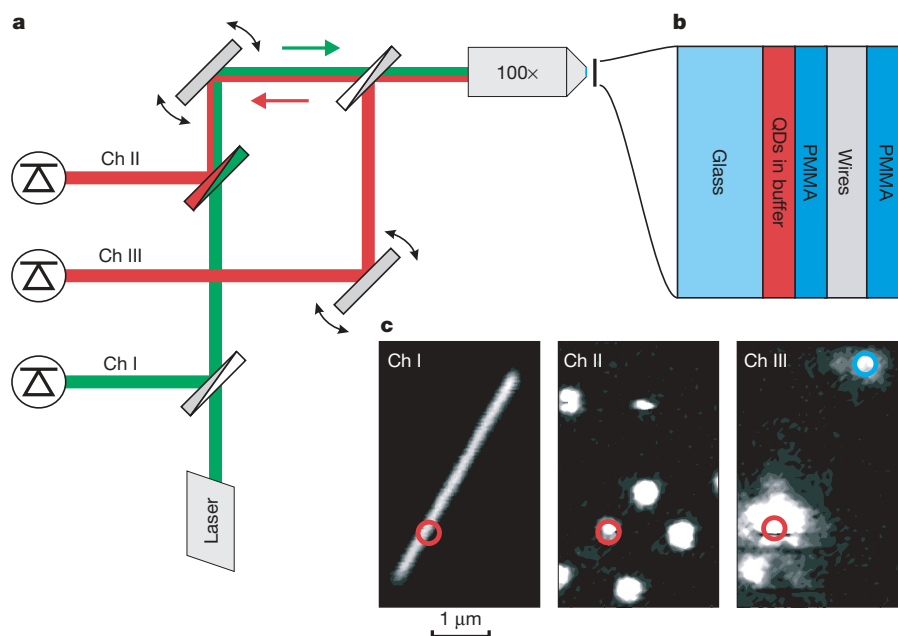
surface plasmon modes localized transversely to dimensions comparable to the wire diameter  $d$ , even when it is much smaller than the optical wavelength  $\lambda$ . This subwavelength localization is accompanied by a dramatic concentration of optical fields<sup>10,11</sup>. In addition, the surface plasmon modes propagate with greatly reduced velocities because they involve the motion of charge-density waves<sup>9,15,16</sup>.

The emission properties of a nanoscale optical emitter can be significantly modified by the proximity of a nanowire that supports surface plasmons. In principle, three distinct decay channels exist. First, direct optical emission into free-space modes is possible, with a rate modified from that of an isolated quantum dot owing to the proximity of the metallic surface<sup>17</sup>. Second, the optical emitter can be damped non-radiatively owing to ohmic losses in the conductor<sup>17</sup>. Last, and most importantly, the tight field confinement and reduced velocity of surface plasmons can cause the nanowire to capture the majority of spontaneous radiation into the guided surface plasmon modes<sup>9</sup>, much like a lens with extraordinarily high numerical aperture. For an optical emitter placed within the evanescent surface plasmon mode tail, the spontaneous emission rate into the surface plasmons<sup>9</sup> is proportional to  $(\lambda/d)^3$ . In contrast, the free-space emission rate can be enhanced by at most a factor of four, whereas non-radiative damping becomes significant only for very small wire-emitter separation<sup>9</sup>. Thus, for an optimally placed emitter, the spontaneous emission rate  $\Gamma_{\text{pl}}$  into surface plasmons can far exceed the radiative and non-radiative rates ( $\Gamma_{\text{rad}}$  and  $\Gamma_{\text{nrd}}$ , respectively), which results in highly efficient coupling to surface plasmons and enhancement of the total decay rate ( $\Gamma_{\text{total}}$ ) compared to that of an uncoupled emitter ( $\Gamma_0$ ). This enhancement can be characterized by a Purcell factor,  $P = \Gamma_{\text{total}}/\Gamma_0$ , which for thin wires is predicted to be large<sup>9</sup>. We emphasize that this strong coupling is caused by the geometrical effect of tight confinement of the surface plasmons, and occurs far away from the plasmon resonance frequency of nanowires<sup>18</sup>. It does not involve an optical cavity<sup>2,3,5–8</sup>, and can be achieved simultaneously over a broad continuum of optical frequencies.

Chemically synthesized CdSe quantum dots<sup>19</sup> placed proximally to silver nanowires comprise a simple experimental system to investigate the emitter–surface plasmon coupling. As illustrated in Fig. 1a, the spontaneous emission of a quantum dot is split between photon

emission into free space, which can be detected by an optical microscope, and the excitation of surface plasmons ( $\Gamma_{\text{nrd}}$  is negligible for our chosen parameters, as described below). During propagation along the smooth nanowire, surface plasmons do not couple to the observable far-field modes of the surrounding dielectric. However, much like a conventional antenna, an abrupt end of the wire can scatter surface plasmons radiatively into far-field modes, thus facilitating their detection using an optical microscope. A simulation of this effect is shown in Fig. 1c, where a quantum dot is placed 25 nm away from one wire end: whereas the surface plasmons decay evanescently away from the nanowire edge, substantial emission into free space results from surface plasmon scattering at the far end of the wire. Silver nanowires were prepared using a solution-phase polyol method with modifications for surface passivation<sup>20</sup> (Supplementary Information). The samples were created by spinning quantum dots onto a glass substrate, covering them with an  $\sim 30$ -nm layer of poly(methylmethacrylate) (PMMA; see Supplementary Information for detailed analysis of the PMMA layer), and then depositing dry wires on top. Finally, the sample was overcoated with a thick layer of PMMA. Scanning electron microscopy images revealed that the diameters of the silver nanowires were  $102 \pm 24$  nm (Supplementary Information). The closest allowed distance between the quantum dots and nanowires is determined by the thickness of the PMMA layer and the quantum dot shell radius ( $\sim 5$  nm), and is  $\sim 35$  nm (Methods and Supplementary Information). The experimental set-up for studying the quantum dot–nanowire system (Fig. 2a) is based on a modified confocal microscope with three scanning channels.

In general, the coupling between an optical emitter and single surface plasmons should be stronger for thinner wires<sup>9</sup> (Fig. 1b). However, for thinner wires, the out-coupling efficiency of surface plasmons to the far-field at the wire end decreases owing to a large wavevector mismatch. In this case, significant surface plasmon reflection at the nanowire ends causes standing surface plasmon wave



**Figure 2 | Experimental set-up.** **a**, Three-channel confocal microscope with 532 nm laser excitation source. **b**, Layout of sample containing quantum dots and nanowires. **c**, Left, channel I: nanowire image. Middle, channel II: image of quantum dots. The red circle denotes the position of the coupled quantum dot, and the same point is also denoted in the leftmost image.

Right, channel III: the excitation laser was focused on the quantum dot (red circle). The largest bright spot corresponds to the quantum dot fluorescence, while two smaller spots correspond to surface plasmons scattered from the nanowire ends. The blue circle indicates the farthest end of the nanowire, used for photon cross-correlation measurements.



formation within the nanowire<sup>12</sup> (Fig. 1c) and eventual energy loss due to heating (ohmic losses). The effect of nanowire diameter on out-coupling efficiency is illustrated in Fig. 1d, where the intensity of the scattered radiation from the wire end is plotted for different wire diameters. For a 25-nm nanowire, hardly any scattering is seen from the end despite the stronger coupling between the emitter and surface plasmons, but the scattering is significant for a 100-nm wire (this was verified experimentally by exciting surface plasmons directly with a laser focused at one wire end; Supplementary Information). Nanowires with  $d \approx 100$  nm exhibit both reasonable emitter–surface plasmon couplings and surface plasmon to far-field scattering, and thus were chosen for the experiments. The large bandwidth of the surface plasmon–emitter coupling enables us to perform the experiments at room temperature, where a single quantum dot spectral width exceeds 15 nm (Supplementary Information).

Figure 2c presents an experimental demonstration of directed emission of a quantum dot into surface plasmons. The leftmost panel shows a confocal reflection image of a silver nanowire recorded with channel I. The middle panel corresponds to a fluorescence image of quantum dots detected at 655 nm with channel II. These two images were used to determine the positions of the nanowire and quantum dot relative to each other. Owing to the resolution limit of our optical system, the actual distance between a quantum dot and the nanowire could not be determined, and only quantum dots that appear directly on top of a nanowire were chosen for experiment. The rightmost panel shows a coupled wire–dot system imaged with channel III. When the proximal quantum dot (circled in red) was excited by the laser, the nanowire ends literally lit up. The large spot around the red circle corresponds to emission from the quantum dot itself, whereas the two other points coincide with the wire ends. Significantly, a high degree of correlation was seen between the time traces of the fluorescence counts from the quantum dot and the end of the coupled wire (Fig. 3a). These observations indicate that the source of the fluorescence from the wire end is the quantum dot.

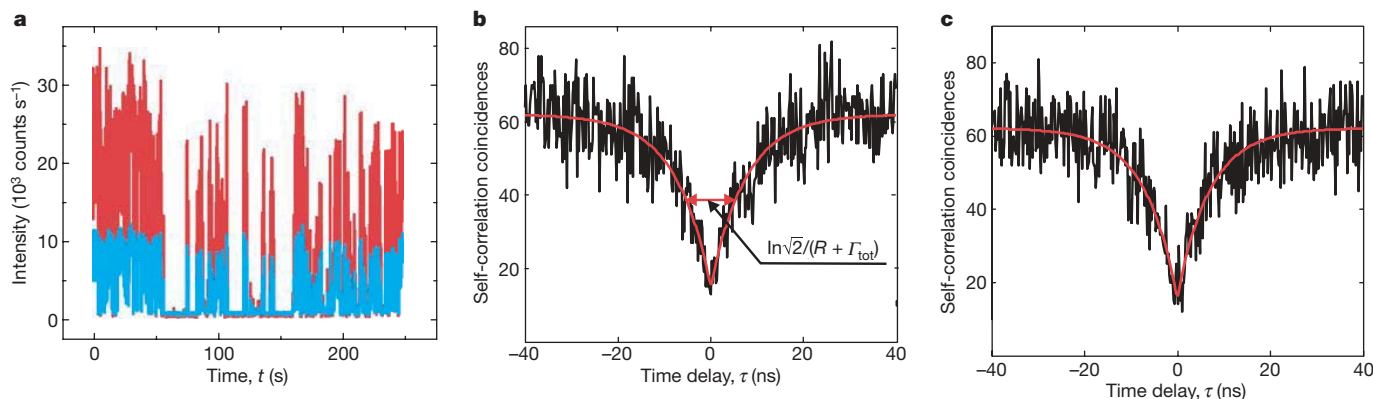
Photon coincidence measurements<sup>1</sup> of the quantum dots (Fig. 3b) demonstrate that these quantum dots can only emit a single photon at a time. In these measurements, the free-space fluorescence from the quantum dot was equally split into two channels using a beam splitter and detected by avalanche photo-diodes. The coincidences between two channels were recorded as a function of time delay  $\tau$ . If the quantum dot emits only one photon at a time it can only be recorded at one of the channels, and therefore zero coincidences are expected between the two channels at  $\tau = 0$ , as seen in Fig. 3b. The slight offset from zero can be attributed to stray light, dark

counts of the detectors and the resolution limit of the electronics (Supplementary Information).

The light emission at the nanowire end is a result of single, quantized surface plasmons scattering off the ends of the nanowire. This is demonstrated in Fig. 3c by the dip at  $\tau = 0$  in the photon coincidence measurements between the free-space fluorescence of the quantum dot and emission from the wire end. This near-zero coincidence is a consequence of the fact that the single photon emitted from a quantum dot can either radiate into free space or the surface plasmon modes, but never both simultaneously.

Data presented in Fig. 3, along with measured count rates, can be used to quantify the coupling strength of the quantum dot to the surface plasmons. As this coupling creates a new decay channel for the quantum dot, its decay rate is expected to increase. To study this enhancement, observed coincidence data were fitted to a simple two-level model of quantum dot emission<sup>21</sup> (Fig. 3b; see also Supplementary Information). The model incorporates an incoherent pumping rate  $R$  from the ground to an excited state of a quantum dot and a decay rate  $\Gamma_{\text{total}}$  back to the ground state. In this model, the temporal width of the anti-bunching dip is given by  $\Delta\tau = \ln\sqrt{2}/(R + \Gamma_{\text{total}})$ , where the excitation rate  $R$  is proportional to the incident power. Therefore, by extracting  $\Delta\tau$  from coincidence measurements as a function of incident laser power and extrapolating to  $R = 0$ ,  $\Gamma_{\text{total}}$  can be obtained (Fig. 4a).

The natural lifetimes of individual dots (20–30 ns) vary owing to the heterogeneity in their structures. However, a comparison of the lifetime distributions of 30 coupled and 100 uncoupled quantum dots (Fig. 4b) clearly demonstrates that statistically the lifetime (decay rate) of the exciton in coupled quantum dots is shortened (enhanced). The average lifetime of the coupled (uncoupled) quantum dots was found to be  $13 \pm 4$  ns ( $22 \pm 5$  ns). At the same time, the distribution for coupled quantum dots has a larger weight towards shorter lifetimes. Specifically, certain coupled and uncoupled quantum dots exhibited lifetimes as short as 6 ns and 15 ns, respectively, indicating that  $P > 2.5$  is achieved for some coupled quantum dot–nanowire systems. The apparent efficiency of emission into the surface plasmons can be estimated by comparing the ratio of photon counts ( $n$ ) obtained directly from the dot and from the wire ends,  $\eta_m \approx n_{\text{ends}}/(n_{\text{dot}} + n_{\text{ends}})$ , and is found to be  $\sim 27\%$  for the best coupled quantum dot–nanowire system (Fig. 4c and Supplementary Information). This value does not account for the surface plasmons that are dissipated before they reach the wire ends. Correcting for the measured average absorption lengths in our nanowires allows us to deduce that the actual efficiency approaches  $\eta \approx 60 \pm 10\%$  (Supplementary Information), directly demonstrating



**Figure 3 | Demonstration of single surface plasmon generation.** **a**, Time trace of fluorescence counts (red curve) from a coupled quantum dot and scattered light (blue) from the end of the coupled nanowire. Fluctuations are due to quantum dot blinking<sup>19</sup>. **b**, Second-order correlation function  $G^{(2)}(\tau)$  of quantum dot fluorescence. The number of coincidences at  $\tau = 0$  goes almost to zero, confirming that the quantum dot is a single-photon source.

The width of the dip depends on  $\Gamma_{\text{total}}$  and the pumping rate  $R$  as shown. **c**, Second-order cross-correlation function between fluorescence of the quantum dot and scattering from the nanowire end, obtained by coincidences between channel II (quantum dot) and channel III (wire end). The black and red traces in **b**, **c** indicate experimental data and best fits, respectively.

very efficient coupling to surface plasmons. We note that this coupling efficiency significantly exceeds that recently observed between atoms and dielectric nanofibres<sup>14,31</sup>.

The broadband nature of the strong coupling is demonstrated by comparing the optical spectra associated with emission from the quantum dot and from the wire end. For individual dots randomly drawn from an inhomogeneous ensemble with  $\lambda = 655 \pm 15$  nm, we find that both the quantum dot and wire-end emission exhibit identical  $\sim 15$ -nm-wide spectra. This is consistent with the ability of metallic wires to guide a broad range of optical frequencies<sup>22</sup> and with theoretical predictions (Supplementary Information) that strong coupling can be obtained for a broad continuum of frequencies away from the peak of the observed plasmon resonances<sup>18</sup>.

Further insight into the quantum dot–surface plasmon coupling can be obtained by comparing our experimental observations with detailed electrodynamic calculations<sup>9</sup>. Our model of quantum dot emission near a silver nanowire embedded in a dielectric medium includes losses as well as multiple surface plasmon modes. Figure 1b shows the total spontaneous emission rates and the efficiency  $\eta = \Gamma_{\text{pl}}/\Gamma_{\text{total}}$  for single surface plasmon generation as a function of quantum dot distance from the wire ( $d = 50$  and  $100$  nm). Here the polarization of the quantum dot transition was selected to be radially oriented, because this direction is expected to yield the dominant contribution to enhancement. For quantum dots positioned  $35$  nm from the wire and for a  $100$  nm wire, the calculation yields a Purcell factor  $P \approx 3.7$ . The lower enhancement observed experimentally can be attributed to the contributions from other polarizations and the random positioning of the quantum dots away from the wire. For this distance of separation, the non-radiative decay rate ( $\Gamma_{\text{nr}} < 0.05\Gamma_0$ ) is predicted to be negligible (Supplementary Fig. 10). In addition to enhanced emission into surface plasmon modes, our theory also predicts a moderate increase in the radiative emission

rate, a well-known phenomenon for dipoles oriented perpendicularly to a metallic surface<sup>17</sup>. For  $100$  nm wires and  $35$  nm nanowire–quantum dot distances, the surface plasmon generation efficiency  $\eta$  is theoretically estimated to be  $\sim 50\%$ , which is consistent with our observations.

Further comparison with theoretical predictions is obtained by repeating our observations with thicker PMMA layers (Fig. 4c, d). These measurements demonstrate that both enhancement and estimated coupling efficiency rapidly decrease as the minimum quantum dot–nanowire spacing increases, and become very small for PMMA thicknesses above  $100$  nm. These observations are also in good agreement with the above theoretical predictions. The large variances in the Purcell factors obtained for different devices are due primarily to variations in the distance between quantum dots and nanowires beyond the minimum allowed distance set by the PMMA layer thickness.

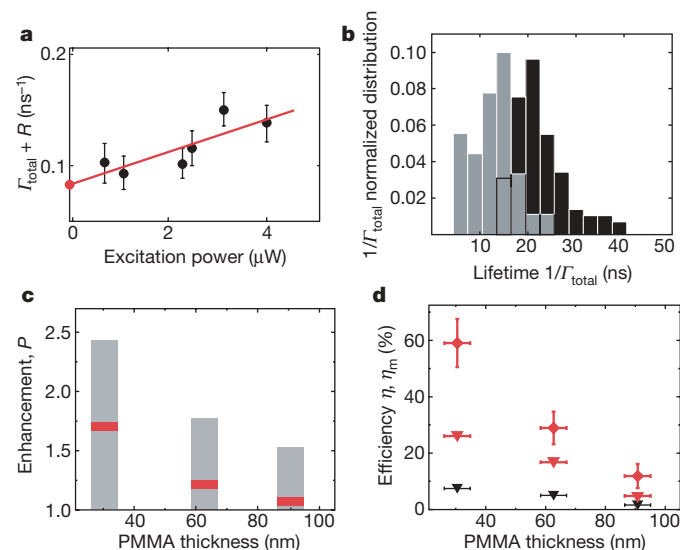
The unique properties of nanoscale surface plasmons have recently been explored in a variety of fascinating systems, from transmission through subwavelength structures<sup>11</sup> to biomedical devices<sup>10</sup> and proposals for realizing ‘perfect’ lenses and invisibility cloaks<sup>10</sup>. Enhancement of fluorescence<sup>23,24</sup>, polarization-dependent coupling<sup>25,26</sup> and normal mode splitting<sup>27,28</sup> near subwavelength structures have also recently been observed. The present work extends these developments in two principal directions. First, we have shown experimentally and theoretically that the present approach results simultaneously in significant enhancement of surface plasmon emission and efficient collection into guided modes propagating along a well-defined direction. Second, it establishes direct coupling between individual emitters and individual, quantized surface plasmons. It thus bridges the fields of nanoscale plasmonics and quantum optics, and opens up the possibility of using quantum optical techniques to achieve new levels of control over the interaction of single surface plasmons and to realize novel quantum plasmonic devices.

In the current set-up, the benefits of using smaller wires must be balanced against poor out-coupling to free-space modes. However, this trade-off can be circumvented by using optimized geometries and evanescent out-coupling to mode-matched optical fibres<sup>9,16,23</sup>. The excellent coupling expected from these integrated systems can be used, for example, for efficient single-photon sources, high resolution microscopy<sup>29,30</sup> and sensing<sup>20</sup>, or long-range quantum bit coupling<sup>3</sup>. Furthermore, in such systems an individual emitter can be made optically opaque to single incident surface plasmons, which can be used to produce large optical nonlinearities for realization of single-photon transistors<sup>4</sup>. Beyond these specific applications, the ability to create and control individual quanta of current oscillating at optical frequencies and accompanied by guided radiation with subwavelength localization opens up intriguing new possibilities at the interface of optics and electronics.

## METHODS SUMMARY

Samples were prepared by spin-coating a solution of chemically synthesized CdSe quantum dots (mixed with  $\text{Na}_2\text{B}_4\text{O}_7$  and cysteine) onto a plasma-cleaned glass slide at  $3,000$  r.p.m. for  $60$  s under a nitrogen atmosphere. Three minutes later, PMMA ( $1$ ,  $2$  and  $3$  wt% in toluene for  $30$ ,  $60$  and  $90$  nm films) was spun on top at  $6,000$  r.p.m. for  $60$  s. The quantum dots used do not dissolve in toluene and are unperturbed during the spin-coating process (experimentally, we find that the arrangement of quantum dots on the surface remains unchanged). A stamp with the modified silver nanowires was placed on top of the slide and pressed for a few seconds. The stamp was left there for  $20$  min and then gently peeled off, leaving nanowires on the PMMA. Finally, PMMA ( $2.2$  wt%) was spun on top at  $1,000$  r.p.m. for  $60$  s (Fig. 2b).

Our confocal microscope uses a c.w.  $532$  nm laser as the excitation source. It is focused onto the sample using a Nikon CFI Plan Fluor  $100\times$  oil immersion objective NA  $1.3$ , while a mirror mounted on a galvanometer is used to scan the incoming beam. Channel II acts as a confocal microscope and is used to image single quantum dots, via fluorescence at  $655$  nm. Channel I is combined with channel II using a  $90:10$  beam splitter that directs part of the reflected laser light towards a detector and can be used to image the silver nanowires. Channel III is combined with the main set-up using a  $50:50$  beam splitter and is an independent



**Figure 4 | Characterization of quantum dot–nanowire coupling.** **a**, The linear dependence of  $G^{(2)}$  width on laser power (black filled circles) is extrapolated to zero power (red filled circle), yielding  $\Gamma_{\text{total}}$  of a quantum dot. The laser power is proportional to the incoherent pumping rate  $R$  of the dot. **b**, Normalized histograms of quantum dot lifetimes. The black (grey) bars denote the distribution of uncoupled (coupled) quantum dots. Overlapping parts of the histograms are indicated by outlined and vertically stacked bars. **c**, Average Purcell enhancement,  $P$ , versus PMMA thickness. Red line, average value of  $P$ . Height and width of grey bars indicate the standard deviations of  $P$  and PMMA thickness, respectively. **d**, Measured maximum and average efficiencies of emission into the surface plasmons versus PMMA thickness. Black (red) filled triangles, average (maximum) apparent coupling efficiencies  $\eta_m$ , without compensating for surface plasmon losses. Red filled diamonds, maximum actual efficiency  $\eta$ , after compensating for dissipation. Error bars in **a**, **c**, **d** indicate  $\pm 1$  s.d.

imaging system. It also includes a galvanometer which allows us to image any diffraction limited spot within the field of view to detect fluorescence at 655 nm. Additional details of our experimental set-up are provided in Supplementary Information.

Received 10 April; accepted 4 September 2007.

1. Haroche, S. & Raimond, J.-M. *Exploring the Quantum: Atoms, Cavities, Photons*. (Oxford Univ. Press, New York, 2006).
2. Englund, D. *et al.* Controlling the spontaneous emission rate of single quantum in a two-dimensional photonic crystal. *Phys. Rev. Lett.* **95**, 013904 (2005).
3. Birnbaum, K. M. *et al.* Photon blockade in an optical cavity with one trapped atom. *Nature* **436**, 87–90 (2005).
4. Chang, D. E., Sørensen, A. S., Demler, E. A. & Lukin, M. D. A single-photon transistor using nano-scale surface plasmons. *Nature Phys.* advance online publication doi:10.1038/nphys708 (26 August 2007).
5. Cirac, J. I., Zoller, P., Kimble, H. J. & Mabuchi, H. Quantum state transfer and entanglement distribution among distant nodes in a quantum network. *Phys. Rev. Lett.* **78**, 3221–3224 (1997).
6. Imamoğlu, A. *et al.* Quantum information processing using quantum dot spins and cavity QED. *Phys. Rev. Lett.* **83**, 4204–4207 (1999).
7. Hennessy, K. *et al.* Quantum nature of a strongly coupled single quantum dot-cavity system. *Nature* **445**, 896–899 (2007).
8. Wilk, T., Webster, S. C., Kuhn, A. & Rempe, G. Single-atom single-photon quantum interface. *Science* **317**, 488–490 (2007).
9. Chang, D. E., Sørensen, A. S., Hemmer, P. R. & Lukin, M. D. Quantum optics with surface plasmons. *Phys. Rev. Lett.* **97**, 053002 (2006).
10. Atwater, H. A. The promise of plasmonics. *Sci. Am.* **296**, 56–63 (2007).
11. Genet, C. & Ebbesen, T. W. Light in tiny holes. *Nature* **445**, 39–46 (2007).
12. Sanders, A. W. *et al.* Observation of plasmon propagation, redirection, and fan-out in silver nanowires. *Nano Lett.* **6**, 1822–1826 (2006).
13. Ditlbacher, H. *et al.* Silver nanowires as surface plasmon resonators. *Phys. Rev. Lett.* **95**, 257403 (2005).
14. Nayak, K. P. *et al.* Optical nanofiber as an efficient tool for manipulating and probing atomic fluorescence. *Opt. Express* **15**, 5431–5438 (2007).
15. Takahara, J., Yamagishi, S., Taki, H., Morimoto, A. & Kobayashi, T. Guiding of a one-dimensional optical beam with nanometer diameter. *Opt. Lett.* **22**, 475–477 (1997).
16. Chang, D. E., Sørensen, A. S., Hemmer, P. R. & Lukin, M. D. Strong coupling of single emitters to surface plasmons. *Phys. Rev. B* **76**, 035420 (2007).
17. Chance, R. R., Prock, A. & Silbey, R. Molecular fluorescence and energy transfer near interfaces. *Adv. Chem. Phys.* **37**, 1–65 (1978).
18. Sun, Y., Gates, B., Mayers, B. & Xia, Y. Crystalline silver nanowires by soft solution processing. *Nano Lett.* **2**, 165–168 (2002).
19. Chung, I., Witkoskie, J. B., Cao, J. & Bawendi, M. G. Description of the fluorescence intensity time trace of collections of CdSe nanocrystal quantum dots based on single quantum dot fluorescence blinking statistics. *Phys. Rev. E* **73**, 011106 (2006).
20. Tao, A. *et al.* Langmuir-Blodgett silver nanowire monolayers for molecular sensing using surface-enhanced Raman spectroscopy. *Nano Lett.* **3**, 1229–1233 (2003).
21. Lounis, B., Bechtel, H. A., Gerion, D., Alivisatos, P. & Moerner, W. E. Photon antibunching in single CdSe/ZnS quantum dot fluorescence. *Chem. Phys. Lett.* **329**, 399–404 (2000).
22. Dickson, R. M. & Lyon, L. A. Unidirectional plasmon propagation in metallic nanowires. *J. Phys. Chem. B* **104**, 6095–6098 (2000).
23. Hochberg, M., Baehr-Jones, T., Walker, C. & Scherer, A. Integrated plasmon and dielectric waveguides. *Opt. Express* **12**, 5481–5486 (2004).
24. Biteen, J. S., Lewis, N. S. & Atwater, H. A. Spectral tuning of plasmon-enhanced silicon quantum dot luminescence. *Appl. Phys. Lett.* **88**, 131109 (2006).
25. Zhang, J., Ye, Y.-H., Wang, X., Rochon, P. & Xiao, M. Coupling between semiconductor quantum dots and two-dimensional surface plasmons. *Phys. Rev. B* **72**, 201306(R) (2005).
26. Mertens, H., Biteen, J. S., Atwater, H. A. & Polman, A. Polarization-selective plasmon-enhanced silicon quantum-dot luminescence. *Nano Lett.* **6**, 2622–2625 (2006).
27. Bellessa, J., Bonnard, C. & Plenat, J. C. Strong coupling between surface plasmons and excitons in an organic semiconductor. *Phys. Rev. Lett.* **93**, 036404 (2004).
28. Dintinger, J., Klein, S., Bustos, F., Barnes, W. L. & Ebbesen, T. W. Strong coupling between surface plasmon-polaritons and organic molecules in subwavelength hole arrays. *Phys. Rev. B* **71**, 035424 (2005).
29. Klimov, V. V., Ducloy, M. & Letokhov, V. S. A model of an apertureless scanning microscope with a prolate nanospheroid as a tip and an excited molecule as an object. *Chem. Phys. Lett.* **358**, 192–198 (2002).
30. Smolyaninov, I. I., Elliott, J., Zayats, A. & Davis, C. C. Far-field optical microscopy with a nanometer-scale resolution based on the in-plane magnification by surface plasmon polaritons. *Phys. Rev. Lett.* **94**, 057401 (2005).
31. Sagué, G., Vetsch, E., Alt, W., Meschede, D. & Rauschenbeutel, A. Cold atom physics using ultra-thin optical fibers: light-induced dipole forces and surface interactions. Preprint at (<http://arxiv.org/quant-ph/0701167>) (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We acknowledge discussions with M. Loncar, J. Doyle, A. Sørensen and M.-H. Yoon, and support from the NSF, DARPA, Harvard-MIT CUA, Harvard CNS, the DTO, the Packard Foundation and Samsung Electronics.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to M.D.L. ([lukin@fas.harvard.edu](mailto:lukin@fas.harvard.edu)) and H.P. ([Hongkun\\_Park@harvard.edu](mailto:Hongkun_Park@harvard.edu)).



# Efficient organic carbon burial in the Bengal fan sustained by the Himalayan erosional system

Valier Galy<sup>1</sup>†, Christian France-Lanord<sup>1</sup>, Olivier Beyssac<sup>2</sup>, Pierre Faure<sup>3</sup>, Hermann Kudrass<sup>4</sup> & Fabien Palhol<sup>1</sup>

Continental erosion controls atmospheric carbon dioxide levels on geological timescales through silicate weathering, riverine transport and subsequent burial of organic carbon in oceanic sediments<sup>1</sup>. The efficiency of organic carbon deposition in sedimentary basins is however limited by the organic carbon load capacity of the sediments and organic carbon oxidation in continental margins. At the global scale, previous studies have suggested that about 70 per cent of riverine organic carbon is returned to the atmosphere, such as in the Amazon basin<sup>2–4</sup>. Here we present a comprehensive organic carbon budget for the Himalayan erosional system, including source rocks, river sediments and marine sediments buried in the Bengal fan. We show that organic carbon export is controlled by sediment properties, and that oxidative loss is negligible during transport and deposition to the ocean. Our results indicate that 70 to 85 per cent of the organic carbon is recent organic matter captured during transport, which serves as a net sink for atmospheric carbon dioxide. The amount of organic carbon deposited in the Bengal basin represents about 10 to 20 per cent of the total terrestrial organic carbon buried in oceanic sediments. High erosion rates in the Himalayas generate high sedimentation rates and low oxygen availability in the Bay of Bengal that sustain the observed extreme organic carbon burial efficiency. Active orogenic systems generate enhanced physical erosion and the resulting organic carbon burial buffers atmospheric carbon dioxide levels, thereby exerting a negative feedback on climate over geological timescales.

Photosynthesis of organic matter is a major pathway for consumption of atmospheric CO<sub>2</sub>. Although most photosynthetic organic carbon (C<sub>org</sub>) is re-oxidized and returns to the atmosphere, a small fraction is buried in sedimentary basins and stored over geological timescales<sup>5</sup>. This burial represents the second largest atmospheric CO<sub>2</sub> sink (after silicate weathering coupled to carbonate precipitation) and contributes to long-term climate regulation<sup>6</sup>. Continental erosion exerts a primary control on C<sub>org</sub> burial through sediment transport and detrital deposition in sedimentary basins. However, assessing the role of continental erosion in this part of the C cycle is complex, as several processes control its efficiency. First, C<sub>org</sub> transported by rivers is composed of both recent organic matter and fossil refractory C<sub>org</sub> derived from erosion of carbonaceous rocks. Erosion-burial of the latter has no effect on the long term C cycle and it is therefore necessary to determine its proportion. Second, it is generally believed that ~70 % of C<sub>org</sub> exported by global rivers (for example, the Amazon) is oxidized in the continental margins before burial and thus returns to the atmosphere<sup>2–4</sup>. Last, assessing the riverine C<sub>org</sub> flux is not straightforward, because the C<sub>org</sub> content of sediment is highly variable owing to transport segregation processes. Understanding the impact of continental erosion on the C<sub>org</sub> cycle

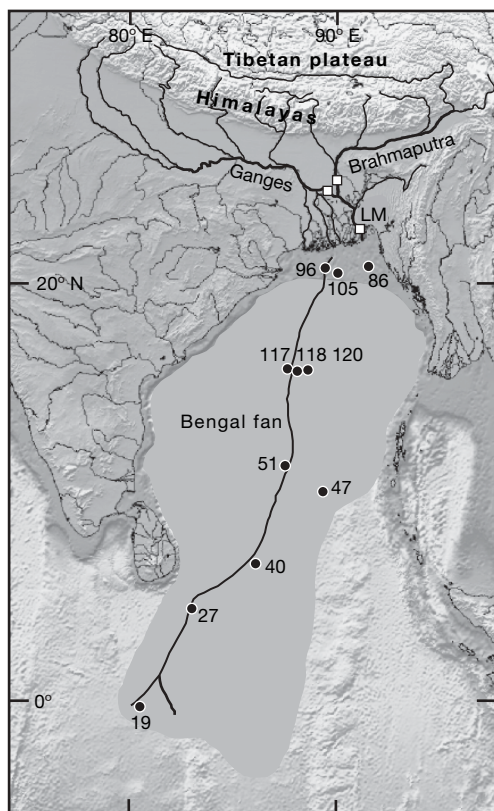
requires identification of the controls on the flux of riverine C<sub>org</sub>, the proportion of rock-derived fossil C<sub>org</sub> and the burial efficiency.

Himalayan erosion generates the largest flux of sediments to the oceans. Today, this represents between 1 and 2 billion tons of sediments exported each year from the Himalayas through the Ganges-Brahmaputra (G-B) system and buried in the Bengal fan sedimentary unit<sup>7–9</sup>. The total C<sub>org</sub> concentration (TOC) in G-B fluvial sediments was hitherto estimated using only surface suspended sediments and without correction for the fossil C<sub>org</sub> contribution<sup>10,11</sup>. Over the past 15 Myr, the Bengal fan has buried an average of about  $0.6 \times 10^{12}$  mol C<sub>org</sub> yr<sup>-1</sup>, that is, 15% of the global burial flux<sup>12</sup>. The Himalayas are thus a key locality for isolating the role of major orogens on the C cycle.

A major sample set covering the whole basin from the Himalayas to the distal part of the Bengal fan has been analysed for C<sub>org</sub>, <sup>14</sup>C and chemistry (Supplementary Tables 1 and 2). The two major rivers (Ganges and Brahmaputra) and their confluence (Lower Meghna) were sampled during three monsoon seasons when ~95% of the sediment flux is exported<sup>13</sup> (Fig. 1, Methods). Bed loads and depth profiles of suspended sediments were collected to integrate the total sediment variability. Simultaneous Acoustic Doppler Current Profiler measurements were performed in order to characterize the hydrodynamic conditions. The Bengal fan sedimentary units are documented by subsurface sediments cored in the shelf, in the mid-fan active channel-levee system and in the distal part of the fan during RV *Sonne* cruise SO93 in 1994<sup>14</sup> (Fig. 1). Detailed study of the modern rivers and their complements in the sedimentary record allows us to quantify: (1) the Himalayan C<sub>org</sub> flux, (2) the proportion of rock-derived fossil C<sub>org</sub> and (3) the terrestrial C<sub>org</sub> burial efficiency.

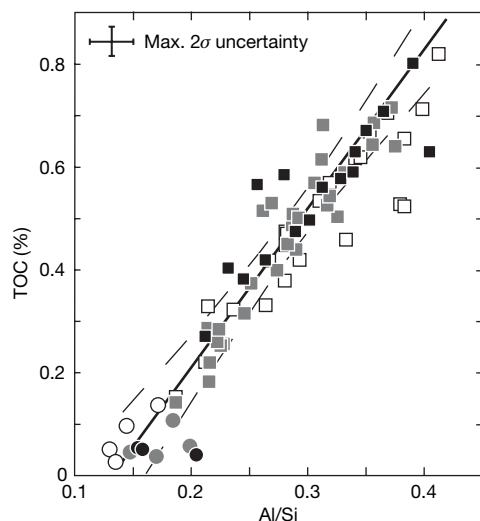
In the G-B system, sediments have a wide range of composition owing to differentiation during transport under hydrodynamic forces. Sediment concentrations increase, while flow velocities decrease, with depth (Supplementary Fig. 1). Suspended sediments are fine-grained and clay-rich in surface water, but become coarser and quartz-rich at depth; bed sediments are sandy and enriched in quartz. TOC varies from 0.03% to 0.82% and is remarkably well correlated with chemical composition; all G-B sediments define a positive correlation between TOC and Al/Si ratio (Fig. 2). This trend mostly expresses co-variation of TOC with grain size and mineralogy, as Al/Si increases with the proportion of aluminous and fine-grained minerals<sup>8</sup>. Individually, the Ganges, the Brahmaputra and the Lower Meghna define comparable trends, indicating identical C<sub>org</sub> loading (Fig. 2). In the G-B system, mineral particles and C<sub>org</sub> appear to have a similar response to hydrodynamic sorting, indicating a tight physical association between these components, either at the grain scale (C<sub>org</sub> sorption on minerals surfaces<sup>15</sup>) or at larger scale (formation of C<sub>org</sub>-mineral aggregates<sup>16</sup>). A consequence of the gradient in TOC is

<sup>1</sup>Nancy Université, CRPG CNRS/INSU, BP 20, 54501 Vandœuvre-lès-Nancy, France. <sup>2</sup>ENS Paris, Laboratoire de Géologie, CNRS/INSU, 24 rue Lhomond, F-75005 Paris Cedex 5, France. <sup>3</sup>Nancy Université, G2R CNRS/INSU, BP 239, 54506 Vandœuvre-lès-Nancy, France. <sup>4</sup>Federal Institute for Geosciences and Natural Resources (BGR), 30655 Hannover, Germany. †Present address: Woods Hole Oceanographic Institution, 360 Woods Hole Road, Woods Hole, Massachusetts 02543, USA.



**Figure 1 | Map of the Himalayan system.** Ganges, Brahmaputra and Lower Meghna (LM) were sampled in Bangladesh (open squares). The different parts of the Bengal fan were investigated using cores from cruise SO 93 (filled black circles and corresponding core numbers). The black line represents the modern active channel.

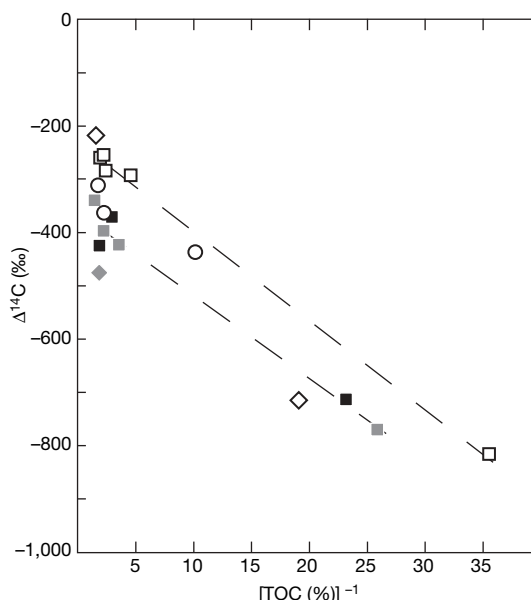
that earlier estimates of  $C_{org}$  fluxes based on surface suspended sediment data were overestimated<sup>10,11</sup>. Thus, assessing the  $C_{org}$  flux requires the mean sediment composition transported by the river



**Figure 2 | TOC of riverine sediments as a function of their Al/Si ratio.** Filled grey symbols, Brahmaputra; open symbols, Ganges; filled black symbols, Lower Meghna. TOC is very low in the sandy and quartz-enriched bed sediments (circles). In the suspended sediments (squares), TOC linearly increases with the relative proportion of aluminous fine-grained minerals. Sediments from the three rivers define comparable positive trends, indicating they have similar  $C_{org}$  loading relative to their composition. Best fit (solid line) and 95% confidence interval (dashed lines) are shown for Lower Meghna sediments.

to be calculated. With our data, the integrated suspended sediment TOC can be estimated by integrating TOC gradient, sediment concentration and flow velocity over the whole river depth. For the Lower Meghna before its discharge into the Bay of Bengal (Fig. 1), we calculate an integrated TOC of  $0.41 \pm 0.04\%$ . In comparison, bed sediments have low TOC with an average value of 0.05%, consistent with their high proportion of quartz.

$C_{org}$  transported by rivers is a mix of inherited fossil  $C_{org}$  ( $FC_{org}$ ) derived from erosion of carbonaceous rocks, and different sources of recent  $C_{org}$  ( $RC_{org}$ ) that include pre-aged soil  $C_{org}$  and fresh plant debris. As  $FC_{org}$  burial has no effect on the long-term atmospheric  $CO_2$  content, its proportion must be determined. We measured  $^{14}C$  concentrations of acid-insoluble  $C_{org}$  composed of  $FC_{org}$  that has no  $^{14}C$  ( $\Delta^{14}C = -1,000\%$ ) and  $RC_{org}$  that is enriched in  $^{14}C$  (ref. 17).  $\Delta^{14}C$  ranges from  $-814\%$  to  $-215\%$  (13.4 to 2.0 kyr) and co-varies with TOC; Ganges, Brahmaputra and Lower Meghna sediments define independent but similar trends. This reflects mixing between bed sediments dominated by  $FC_{org}$  and surface suspended sediments dominated by  $RC_{org}$  (Fig. 3). The  $RC_{org}$   $\Delta^{14}C$  is unknown, as it is derived from sources of variable ages. Surface suspended sediments constrain minimum  $\Delta^{14}C$  values for the  $RC_{org}$  of each river.  $\Delta^{14}C$  of  $RC_{org}$  must be higher, as  $FC_{org}$  has been detected in these surface sediments by optical and Raman spectroscopy methods (Supplementary Figs 3 and 4, Supplementary Discussion). Therefore, the highest measured  $\Delta^{14}C$  sets the minimum  $FC_{org}$  to 14% in average suspended sediment and 42% in bed sediment. On the basis of Himalayan river gravel and bedrock<sup>10,12,18</sup> data (Supplementary Table 1, Supplementary Fig. 2), average  $C_{org}$  content in the source rocks is between 0.05% and 0.08%. Using the TOC versus Al/Si relationship (Fig. 2) and assuming there is no fossil  $C_{org}$  oxidation during erosion and transport, we can limit the maximum proportion of  $FC_{org}$  to 20% in average suspended sediment and to 100% in bed sediment. Combining the  $C_{org}$  content in source rocks with  $\Delta^{14}C$  constraints, we conservatively estimate that  $FC_{org}$  represents 10–20% of TOC in suspended sediments and 40–100% of TOC in bed sediments.

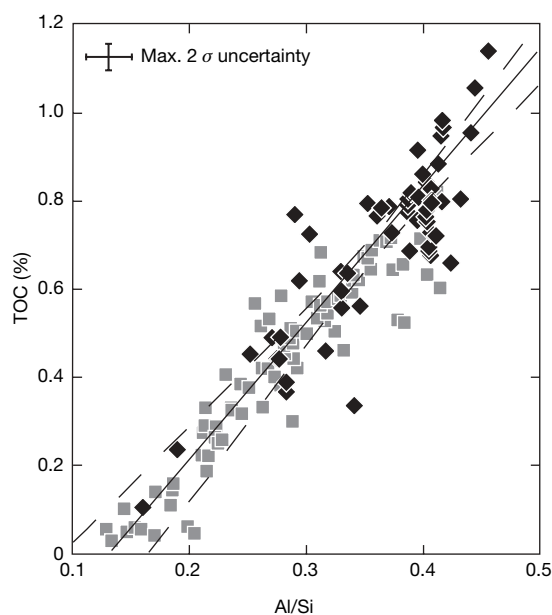


**Figure 3 | Radiocarbon composition of acid-insoluble  $C_{org}$  extracted from riverine depth profile sediments as a function of  $1/[TOC (\%)]^{-1}$ .** Filled grey symbols, Brahmaputra; open symbols, Ganges; and filled black symbols, Lower Meghna. Circles, monsoon 2002; squares, monsoon 2004; diamonds, monsoon 2005. Sediments from each river define a mixing trend between bed sediment dominated by aged  $C_{org}$  and suspended sediments with higher TOC and dominated by modern  $C_{org}$ . Maximum  $2\sigma$  uncertainties on  $\Delta^{14}C$  and TOC are respectively 5‰ and 0.03%.

Suspended sediment fluxes have been measured for the Ganges and the Brahmaputra, and the average total flux is around  $1 \times 10^9 \text{ t yr}^{-1}$  (ref. 13). The flux of bed sediment is not directly measured, but geochemical mass balance implies that bed sediment flux plus floodplain sequestration are almost equal to the suspended sediment flux<sup>8</sup>. Taking into account these sediment fluxes, we estimate that the G-B system delivers  $(3.1 \pm 0.3) \times 10^{11} \text{ mol yr}^{-1}$  of  $\text{RC}_{\text{org}}$  to the Bay of Bengal.

Quantifying the corresponding  $\text{CO}_2$  uptake requires determination of the proportion of this  $\text{RC}_{\text{org}}$  that is preserved and effectively buried in the Bengal fan. Sediment delivery to the Bengal delta shelf and active channel-levee system<sup>14</sup>, which directly reflects the input of the G-B and their sources, has been stable over the past 20 kyr, as shown by constant Sr and Nd isotopic ratios<sup>19</sup>. In these sediments, TOC and Al/Si define a positive linear correlation statistically identical to that defined by G-B sediments (Fig. 4). It indicates similar  $\text{C}_{\text{org}}$  loads for river and marine sediments. The relative abundances of terrestrial and marine biomarkers as well as their stable isotopic compositions show that  $\text{C}_{\text{org}}$  is largely dominated by terrestrial matter (Supplementary Table 3, Supplementary Fig. 5, Supplementary Discussion), as already observed for ancient Bengal fan sediments<sup>20–22</sup>. Therefore, the terrestrial  $\text{C}_{\text{org}}$  loading of Bengal fan sediments roughly equals that of G-B river sediments. The Bengal fan appears to be able to preserve and bury terrestrial  $\text{C}_{\text{org}}$  that is exported by the G-B system. This strongly contrasts with other active margins where the burial efficiency is much lower and limited to ~30% (refs 2–4).

This extreme burial efficiency probably derives from the transport dynamic, with rapid sediment accumulation on the delta shelf (up to  $30 \text{ cm yr}^{-1}$ ; ref. 23). In shallow water, where  $\text{C}_{\text{org}}$  oxidation is theoretically most efficient,  $\text{C}_{\text{org}}$  is protected from oxidation by quick burial under a thick layer of fresh sediments. Typical  $\text{O}_2$  penetration depth in such sediments is only a few centimetres<sup>24</sup>, that is, the same order of magnitude as the Bengal shelf annual accumulation<sup>23</sup>. Therefore,  $\text{C}_{\text{org}}$  exposure time to  $\text{O}_2$  is only a few years or so, insufficient for effective  $\text{C}_{\text{org}}$  oxidation. Sediments are finally transferred to the deep-water zone by turbiditic currents and deposited in channel-levee systems<sup>25,26</sup>. Identical  $\text{C}_{\text{org}}$  loading in the shelf, active channel-levee and mid-deep fan sediments (Supplementary Fig. 6)



**Figure 4 | TOC of riverine and fan sediments as a function of their Al/Si ratio.** Filled black diamonds, TOC of the Bengal fan sediments, with best fit (solid line) and 95% confidence interval (dashed lines). Marine sediments define a positive trend highly comparable to that defined by river sediments (filled grey squares), suggesting they have similar  $\text{C}_{\text{org}}$  loads.

indicates that  $\text{C}_{\text{org}}$  oxidation during sediment transfer from shelf to deep ocean is negligible. In addition, river discharge and precipitation are responsible for a negative salinity anomaly, and maintain stratified waters in the Bay of Bengal<sup>27,28</sup>. Combined with high productivity sustained by high nutrient flux delivered by the G-B, this generates an extended  $\text{O}_2$  minimum zone and generally low  $\text{O}_2$  concentrations in the Bay of Bengal<sup>27,28</sup>.  $\text{C}_{\text{org}}$  oxidation is thus limited by both short exposure to, and low availability of,  $\text{O}_2$ .

Climatic and tectonic factors produce high physical erosion rates and a major influx of particles and nutrients to the ocean that favour long-term  $\text{C}_{\text{org}}$  burial. This is characteristic of the Himalayan basin and is also true for other comparable basins, such as Oceania, that generate comparable sediment influx with high burial efficiency<sup>9,29</sup>. Uplift and erosion result in rapid sediment accumulation, which favours  $\text{C}_{\text{org}}$  preservation with globally significant fluxes. Active orogenesis affects the export and burial of  $\text{C}_{\text{org}}$  for periods of  $10^6$ – $10^7$  years and tends to buffer long-term atmospheric  $\text{CO}_2$  and  $\text{O}_2$  levels.

## METHODS SUMMARY

**Sampling.** Detailed sampling of the Ganges, Brahmaputra and Lower Meghna was performed during the 2002, 2004 and 2005 monsoons in Bangladesh. Suspended sediments were taken at different depths in the river channel and bedload was dredged from the bottom of the river. Marine sediments were taken from short cores drilled in the Bengal shelf, channel-levee system and deep-sea fan by the RV *Sonne* in 1994 (cruise SO 93).

**Organic carbon analyses.** Before organic carbon content and isotopic measurements, sediments were leached with HCl (4 wt%) at  $80^\circ\text{C}$  in order to remove any trace of carbonate. TOC and  $\delta^{13}\text{C}$  of bulk  $\text{C}_{\text{org}}$  were measured by elemental analyser-isotope ratio mass spectrometer (EA-IRMS) on a modified EuroVector EuroEA3028-HT coupled with a GV Isoprime. The overall  $2\sigma$  uncertainties associated with the TOC and  $\delta^{13}\text{C}$  of bulk  $\text{C}_{\text{org}}$  determination are 0.03% and 0.25%, respectively.  $^{14}\text{C}$  measurements were performed on the LMC14 (Gif sur Yvette) AMS national facility. After soluble  $\text{C}_{\text{org}}$  extraction with dichloromethane, n-alkane relative abundances were determined on a HP 5890 gas chromatograph-mass spectrometer (GC-MS) and compound specific  $\delta^{13}\text{C}$  were measured on a GV-Isoprime gas chromatograph-isotope ratio mass spectrometer (GC-IRMS).

**Fossil organic carbon detection.** Fossil organic carbon was identified and characterized by Raman microspectroscopy and high-resolution transmission electron microscopy (HRTEM). Raman spectra were obtained on untreated raw sediments with a Renishaw INVIA Reflex Raman micro-spectrometer at the Laboratoire de Géologie de l'Ecole Normale Supérieure, Paris, France. HRTEM was performed on a JEOL 2010 microscope with an acceleration voltage of 200 kV, after concentration of carbonaceous material by HCl/HF acid treatment at low temperature (below  $70^\circ\text{C}$ ).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 15 February; accepted 11 September 2007.

- Hayes, J. M., Strauss, H. & Kaufman, A. J. The abundance of  $^{13}\text{C}$  in marine organic matter and isotopic fractionation in the global biogeochemical cycle of carbon during the past 800 Ma. *Chem. Geol.* **161**, 103–125 (1999).
- Burdige, D. J. Burial of terrestrial organic matter in marine sediments: A re-assessment. *Glob. Biogeochem. Cycles* **19**, GB4011 1–7 (2005).
- Hedges, J. I., Keil, R. G. & Benner, R. What happens to terrestrial organic matter in the ocean? *Org. Geochem.* **27**, 195–212 (1997).
- Schlünz, B. & Schneider, R. R. Transport of terrestrial organic carbon to the oceans by rivers: Re-estimating flux and burial rates. *Int. J. Earth Sci.* **88**, 599–606 (2000).
- Mackenzie, F. T. in *Flux of Organic Carbon by Rivers to the Ocean* (eds Likens, G. E. et al.) 360–384 (US Department of Energy, Washington DC, 1981).
- Berner, R. A. Atmospheric  $\text{CO}_2$  levels over Phanerozoic time. *Science* **249**, 1382–1386 (1990).
- Curry, J. R., Emmel, F. J. & Moore, D. G. The Bengal Fan: morphology, geometry, stratigraphy, history and processes. *Mar. Petrol. Geol.* **19**, 1191–1223 (2003).
- Galy, A. & France-Lanord, C. Higher erosion rates in the Himalaya: Geochemical constraints on riverine fluxes. *Geology* **29**, 23–26 (2001).
- Milliman, J. D. & Syvitski, P. M. Geomorphic/tectonic control of sediment discharge to the ocean: The importance of small mountainous rivers. *J. Geol.* **100**, 525–544 (1992).
- Aucour, A.-M., France-Lanord, C., Pédajo, K., Pierson-Wickmann, A.-C. & Sheppard, S. M. F. Fluxes and sources of particulate organic carbon in the Ganga-Brahmaputra river system. *Glob. Biogeochem. Cycles* **20**, GB2006 1–12 (2006).



11. Subramanian, V. & Ittekkot, V. in *Biogeochemistry of Major World Rivers* (eds Degens, E. T., Kempe, S. & Richey, J. E.) 157–168 (SCOPE, Paris, 1991).
12. France-Lanord, C. & Derry, L. A. Organic carbon burial forcing of the carbon cycle from Himalayan erosion. *Nature* **390**, 65–67 (1997).
13. RSP. *Spatial Representation and Analysis of Hydraulic and Morphological Data* (Report No. FAP 24, WARPO, Dhaka, Bangladesh, 1996).
14. Ittekkot, V., Kudrass, H. R., Quadfasel, D. & Unger, D. Bay of Bengal. *Deep-sea Res. II* **50**, 853–1053 (2003).
15. Hedges, J. I. & Keil, R. G. Sedimentary organic matter preservation: An assessment and speculative synthesis. *Mar. Chem.* **49**, 81–115 (1995).
16. Wagai, R. & Mayer, L. M. Sorptive stabilization of organic matter in soils by hydrous iron oxides. *Geochim. Cosmochim. Acta* **71**, 25–35 (2007).
17. Drenzek, N. J., Montlucon, D. B., Yunker, M. B., Macdonald, R. W. & Eglinton, T. I. Constraints on the origin of sedimentary organic carbon in the Beaufort Sea from coupled molecular  $^{13}\text{C}$  and  $^{14}\text{C}$  measurements. *Mar. Chem.* **103**, 146–162 (2007).
18. Beck, R. A., Burbank, D. W., Sercombe, W. J., Olson, T. L. & Khan, A. M. Organic carbon exhumation and global warming during the early Himalayan collision. *Geology* **23**, 387–390 (1995).
19. Pierson-Wickmann, A.-C., Reisberg, L., France-Lanord, C. & Kudrass, H. Os-Sr-Nd results from sediments in the Bay of Bengal: Implications for sediment transport and the marine Os record. *Paleoceanography* **16**, 435–444 (2001).
20. France-Lanord, C. & Derry, L. A.  $\delta^{13}\text{C}$  of organic carbon in the Bengal Fan: Source evolution and transport of C3 and C4 plant carbon to marine sediments. *Geochim. Cosmochim. Acta* **58**, 4809–4814 (1994).
21. Freeman, K. H. & Colarusso, L. A. Molecular and isotopic records of C4 grassland expansion in the late Miocene. *Geochim. Cosmochim. Acta* **65**, 1439–1454 (2001).
22. Meyers, P. A. & Dickens, G. R. in *Synthesis of Results from Scientific Drilling in the Indian Ocean* (eds Duncan, R. A. et al.) 295–309 (Vol. 70, Geophysical Monograph Series, American Geophysical Union, Washington, 1992).
23. Suckow, A., Morgenstern, U. & Kudrass, H.-R. Absolute dating of recent sediments in the cyclone-influenced shelf area off Bangladesh: Comparison of gamma spectrometric ( $^{137}\text{Cs}$ ,  $^{210}\text{Pb}$ ,  $^{228}\text{Ra}$ ), radiocarbon, and  $^{32}\text{Si}$  ages. *Radiocarbon* **43**, 917–927 (2001).
24. Cai, W.-J. & Sayles, F. L. Oxygen penetration depths and fluxes in marine sediments. *Mar. Chem.* **52**, 123–131 (1996).
25. Schwenk, T., Spiess, V., Breitzke, M. & Hubscher, C. The architecture and evolution of the Middle Bengal Fan in vicinity of the active channel-levee system imaged by high-resolution seismic data. *Mar. Petrol. Geol.* **22**, 637–656 (2005).
26. Weber, M. E., Wiedicke, M. H., Kudrass, H. R., Hübscher, C. & Erlenkeuser, H. Active growth of the Bengal Fan during sea-level rise and highstand. *Geology* **25**, 315–318 (1997).
27. Berner, U., Poggenburg, J., Faber, E., Quadfasel, D. & Frische, A. Methane in ocean waters of the Bay of Bengal: Its sources and exchange with the atmosphere. *Deep-sea Res. II* **50**, 925–950 (2003).
28. Broecker, W. S., Toggweiler, J. R. & Takahashi, T. The Bay of Bengal — A major nutrient source for the deep Indian Ocean. *Earth Planet. Sci. Lett.* **49**, 506–512 (1980).
29. Goni, M. A. et al. Distribution and sources of particulate organic matter in the water column and sediments of the Fly River Delta, Gulf of Papua (Papua New Guinea). *Estuar. Coast. Shelf Sci.* **69**, 225–245 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This study was funded by CNRS-INSU programmes Eclipse and Relief de la Terre. We thank A. Gajurel and M. Rahman for help during the sampling of river sediments; C. Guilmette for technical assistance in the stable isotopes laboratory; and P. Burnard and L. Reisberg for improving the quality of the text.

**Author Contributions** V.G. and C.F.L. are the main authors. O.B. conducted the fossil  $\text{C}_{\text{org}}$  characterization. P.F. and F.P. were involved in the organic geochemistry part of the study. H.K. provided the Bengal fan sediments and the framework of sedimentological interpretation.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to V.G. ([vgaly@whoi.edu](mailto:vgaly@whoi.edu) or [vgaly@crpg.cnrs-nancy.fr](mailto:vgaly@crpg.cnrs-nancy.fr)).

## METHODS

**Sampling.** Detailed sampling of the Ganges, Brahmaputra and Lower Meghna was performed during the 2002, 2004 and 2005 monsoons in Bangladesh. On the basis of daily river survey data from the Bangladesh Water Development Board (BWDB), sediment flux during the monsoon is three orders of magnitude higher than during the dry season<sup>13</sup>. As a consequence, ~95% of the sediment flux is transported during the monsoon, and suspended sediments sampled during the monsoon are therefore representative of the total sediment flux. Suspended sediments were taken at different depths in the river channel and bedload was dredged from the bottom of the river. The sampling was systematically performed in the main channel where the velocities are the highest. In 2005, we also performed a depth sampling in a low velocity zone close to the bank to improve the description of the whole river section. Sediments were sampled using a 5-l Nalgene bottle, ballasted and equipped with a depth probe. At a given depth the bottle was opened, then lifted on board after filling. During filling, the depth slightly increases by about one metre; the average depth was recorded as the reference depth. The total contents of the bottle were carefully transferred into opaque plastic containers and sealed free of air. Samples were then filtered at 0.2 µm through a 90 mm nitrile membrane. Special care was taken to recover particles adsorbed at the surface of the container. Suspended sediment concentrations were weighed for each sample. Major and trace element concentrations were measured by ICP-AES and ICP-MS at the Service d'Analyse des Roches et des Minéraux (CRPG-Nancy) on bulk sediment after lithium meta-borate fusion.

Marine sediments were taken from short cores drilled in the Bengal fan by the RV *Sonne* in 1994 (cruise SO 93)<sup>14,30</sup>. For cores from the channel-levee system, deposition ages have been estimated on the basis of foraminifer <sup>14</sup>C ages and stable isotopes; sediments used in this study were deposited during the last glacial–interglacial cycle. Sediments from the Bengal shelf were dated using gamma-spectrometric and radiocarbon methods<sup>23</sup>. We used a sample set that covers the past 500 yr of sedimentation, including modern samples.

**Organic carbon analyses.** G-B river sediments as well as Bengal fan sediments contain significant amounts of detrital carbonates including dolomite<sup>31–33</sup>. C<sub>org</sub> concentration and isotopic measurements must therefore be performed on decarbonated sediments. Efficient dolomite dissolution was achieved through 1 h leaching with 4 wt% HCl at 80 °C. After 50 °C oven drying, the C<sub>org</sub> content was determined either by elemental analysis on a EuroVector EuroEA3028-HT or by conventional CuO combustion<sup>20</sup>. The problem with aqueous acid leaching is that an unknown proportion of C<sub>org</sub> is solubilized by acid hydrolysis. For a representative set of river and marine sediments, the amount of C<sub>org</sub> solubilized during the decarbonation was measured in order to obtain the real TOC. C<sub>org</sub> content measured after decarbonation and TOC define a very good linear correlation (Supplementary Fig. 6). We used this relationship to calculate TOC for each sample from C<sub>org</sub> content measured after decarbonation. The overall 2σ uncertainty associated with the TOC determination is 0.03%.

<sup>14</sup>C were also determined on decarbonated fractions. Acid-insoluble C<sub>org</sub> was converted into pure CO<sub>2</sub> at CRPG (Nancy) following the conventional CuO combustion method<sup>20</sup>. <sup>14</sup>C measurements were performed on the LMC14 (Gif

sur Yvette) AMS national facility. Maximum overall uncertainty on <sup>14</sup>C calculations was estimated to be 5%.

Soluble C<sub>org</sub> was extracted from bulk sediments with dichloromethane using an Accelerated Solvent Extractor (ASE 200). n-alkane relative abundances were determined on a HP 5890 GC-MS and compound specific <sup>13</sup>C was measured on a GV-Isoprime GC-IRMS.

**Raman spectroscopy and HRTEM.** Raman spectra were obtained with a Renishaw INVIA Reflex Raman micro-spectrometer at the Laboratoire de Géologie of the Ecole Normale Supérieure, Paris, France. Spectra were excited at room temperature with the 514.5 nm line of a 20 mW Ar Spectra Physics laser through a LEICA 50X long-working distance objective (NA 0.50). The laser spot at the sample surface had a diameter of approximately 3 µm and a power of 1 mW, which should be low enough to avoid any spectral change or sample destruction due to light absorption and local temperature increase<sup>34</sup>. Light was dispersed by a holographic grating with 1,800 grooves mm<sup>-1</sup>. The spectral resolution of about 1.4 cm<sup>-1</sup> was determined by measuring a neon lamp emission. The spectrometer was calibrated for every session by measuring the position of the neon lamp emission and/or a silicon wafer. The dispersed light was collected by a RENCAM CCD detector. The synchroscan mode from 100 to 2,000 cm<sup>-1</sup> was selected in order to retrieve Raman modes from both carbonaceous matter (CM; 1,300–1,600 cm<sup>-1</sup> region) and silicates/carbonates (100–1,100 cm<sup>-1</sup>). Raman investigations were directly performed on the raw samples, without any preparation. The first-order Raman spectrum of disordered CM exhibits a graphite G band at 1,580 cm<sup>-1</sup>, corresponding to in-plane vibration of aromatic carbons, and several defect bands (including the most important at 1,350 cm<sup>-1</sup>), corresponding to physico-chemical defects<sup>34</sup>.

HRTEM was performed on a JEOL 2010 microscope with an acceleration voltage of 200 kV. Low- and high-magnification observations as well as electron diffraction were conducted on river and oceanic sediment samples. In particular, the high-magnification 002 lattice fringes mode was used to image directly the aromatic skeleton of CM together with selected area electron diffraction to detect the presence of triperiodic graphite. For HRTEM investigations, CM was concentrated from river and marine sediments using HCl/HF acid treatment at low temperature (always below 70 °C). More details regarding the sample preparation and transmission electron microscopy analysis of natural solid CM may be found in ref. 35 and references therein.

30. Kudrass, H. *Bengal Fan* 45–89 (Sonne cruise 93 final report, BGR, Hannover, Germany, 1996).
31. Cochran, J. R. *et al.* *Proc. ODP Init. Rep.* **116**, 388 (1989).
32. Galy, A., France-Lanord, C. & Derry, L. A. The strontium isotopic budget of Himalayan rivers in Nepal and Bangladesh. *Geochim. Cosmochim. Acta* **63**, 1905–1925 (1999).
33. von der Borch, C. C. *et al.* *Initial Report of the Deep Sea Drilling Project* Vol. 22 (US Government Printing Office, Washington, 1974).
34. Beyssac, O. *et al.* On the characterisation of disordered and heterogenous carbonaceous materials by Raman spectroscopy. *Spectrochim. Acta A* **59**, 2267–2276 (2003).
35. Beyssac, O., Rouzaud, J.-N., Goffé, B., Brunet, F. & Chopin, C. Graphitization in a high-pressure, low temperature gradient: A Raman microspectroscopy and HRTEM study. *Contrib. Mineral. Petrol.* **143**, 19–31 (2002).

# Cooperation and conflict in quorum-sensing bacterial populations

Stephen P. Diggle<sup>1</sup>, Ashleigh S. Griffin<sup>2</sup>, Genevieve S. Campbell<sup>1</sup> & Stuart A. West<sup>2</sup>

It has been suggested that bacterial cells communicate by releasing and sensing small diffusible signal molecules in a process commonly known as quorum sensing (QS)<sup>1–4</sup>. It is generally assumed that QS is used to coordinate cooperative behaviours at the population level<sup>3,5</sup>. However, evolutionary theory predicts that individuals who communicate and cooperate can be exploited<sup>6–10</sup>. Here we examine the social evolution of QS experimentally in the opportunistic pathogen *Pseudomonas aeruginosa*, and show that although QS can provide a benefit at the group level, exploitative individuals can avoid the cost of producing the QS signal or of performing the cooperative behaviour that is coordinated by QS, and can therefore spread. We also show that a solution to the problem of exploitation is kin selection, if interacting bacterial cells tend to be close relatives. These results show that the problem of exploitation, which has been the focus of considerable attention in animal communication, also arises in bacteria.

An exponentially increasing body of research suggests that communication between bacterial cells can occur by QS. The uptake of QS signalling molecules has two consequences. First, they regulate a variety of behaviours such as the production and secretion of diverse exoproducts, which have various uses: virulence factors that damage the host, nutrient scavenging molecules (for example siderophores), compounds for providing structure for growth in biofilms, and surfactants for facilitating movement. Most of these exofactors seem to provide a benefit to surrounding cells in the local population and represent ‘public goods’ or cooperative behaviours<sup>11,12</sup>. Second, the uptake of signal molecules also leads to an increase in production of the signal molecules (termed autoinduction). This often leads to positive feedback at high cell densities and to a marked increase in the cooperative production of exoproducts<sup>4</sup>. The assumption here is that many cooperative behaviours, such as the production of exoproducts, are not worth performing until a sufficient number of cells (a quorum) are present, and that QS provides a mechanism for stimulating such behaviours only when high enough cell population densities are reached.

It is frequently assumed that QS has been selected for because it optimizes population growth or survival<sup>3,5</sup>. However, evolutionary theory shows that a behaviour can be favoured by selection only if it increases the inclusive fitness of the individual actor, which is often not the same as maximizing population fitness<sup>7–10,12</sup>. In particular, a large body of evolutionary theory, and empirical studies on a range of other organisms, have shown that cooperation and communication can generate conflicts of interest between individuals<sup>6–9</sup>. QS seems to involve using cooperative, honest communication to coordinate other cooperative behaviours, and so a double evolutionary problem arises. Specifically, populations are at risk from the invasion of selfish cheats, who exploit either the signalling or the exoproduct production of others<sup>1,2,13,14</sup>. A cheat that does not produce QS molecules can

benefit from monitoring the local cell density without investing effort into the dissemination of this information (‘signal negative’). Alternatively, a cheat could produce the signal but not increase the production of exoproducts in response (‘signal blind’), exploiting (and encouraging) the cooperative production of exoproducts by other cells.

Here we examine one of the best-understood QS systems, that of *Pseudomonas aeruginosa*, a Gram-negative bacterium, capable of causing disease in plants and animals, including humans<sup>15</sup>. QS is crucial to the success and virulence of *P. aeruginosa* because it controls behaviours such as biofilm development, swarming motility and the production of an arsenal of extracellular virulence factors that are capable of causing extensive tissue damage, bloodstream invasion and consequently promoting systemic dissemination<sup>4,15</sup>. In this bacterium, QS is controlled by two pathways that regulate the production of *N*-acyl homoserine lactone (AHL) signalling molecules (see Supplementary Fig. 1). These two systems are termed *las* and *rhl*, and use different AHL signal molecules, synthesized by means of the signal generators LasI (*N*-(3-oxododecanoyl)-L-homoserine lactone; 3O-C12-HSL) and RhII (*N*-butanoyl-L-homoserine lactone; C4-HSL), respectively<sup>15</sup>. The importance of QS in controlling the behaviours of *P. aeruginosa* is emphasized by the fact that 6–10% of its genome is regulated by the *las* and *rhl* systems<sup>16</sup>.

We examine the fitness consequences of QS and determine the possibilities for cheating. We focus on the *las* QS pathway because this system controls the *rhl* system hierarchically; a mutation in *las* QS results in a general abolition of QS in *P. aeruginosa*<sup>15,17,18</sup>. To examine the costs and benefits of QS, we used two types of cheating QS mutant. First, we used a ‘signal-negative’ mutant that does not produce the 3O-C12-HSL signal but still responds to signal—the *lasI* mutant (PAO1 *lecA::luxΔlasI*<sup>19</sup>). Second, we used a ‘signal-blind’ mutant that does not respond to extracellular signal—the *lasR* mutant (PAO1 *lecA::luxΔlasR*<sup>19</sup>). The signal-blind mutant does not increase signal production in response to extracellular signal and so a relatively small amount of signal is produced. Both mutants are found in natural populations, with the signal-blind strain being more common<sup>20</sup>. Furthermore, we experimentally altered the level of signal perceived by either the wild type or the signal-negative mutant by adding to cultures a synthetic signal chemically identical to that produced by *P. aeruginosa*<sup>21</sup>.

We first examined the fitness consequences of QS in an environment in which cooperation is favoured. Proteases are a group of exoproducts whose production is controlled by QS. *P. aeruginosa* produces a number of well-known proteases such as alkaline protease, the LasA protease, and the LasB protease more commonly known as elastase, which is completely dependent on a functional *las* QS system<sup>15</sup>. In a host, elastase is released from a bacterial cell to digest several proteins including elastin, fibrin and collagen. Digested

<sup>1</sup>Institute of Infection, Immunity & Inflammation, Centre for Biomolecular Sciences, University Park, University of Nottingham, Nottingham NG7 2RD, UK. <sup>2</sup>Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, King’s Buildings, Edinburgh EH9 3JT, UK.



protein can then be internalized by a bacterial cell and used as nutrition to promote growth. We examined the growth of the wild type and the *lasI* (signal-negative) and *lasR* (signal-blind) mutants in a medium in which the ability to make proteases is required for growth; we term this the quorum-sensing growth medium (QSM).

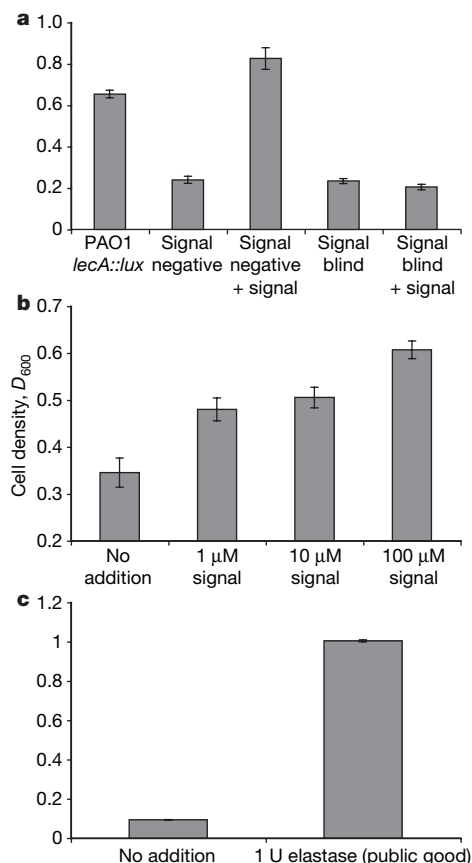
We found that QS provides a benefit at the population level by increasing the production of cooperative exoproducts that can aid growth in certain environmental conditions. Both the signal-negative and the signal-blind mutants showed significantly poorer growth than the parental wild-type strain in the QSM (Fig. 1a;  $F_{(1,32)} = 710.92$ ,  $P < 0.0001$ ). To confirm that this difference in growth yield was caused by QS, we examined the roles of signalling and the production of public goods (protease) separately. Specifically, we found the following: first, the addition of synthetic signal (20  $\mu$ M 3O-C12-HSL) to the signal-negative strain significantly improved growth yield (Fig. 1a;  $F_{(1,31)} = 9.57$ ,  $P < 0.01$ ), as would be expected because this will stimulate the production of proteases such as elastase; second, the signal-negative mutant grew more rapidly as higher concentrations of synthetic signal were added (Fig. 1b;  $F_{(1,22)} = 79.90$ ,  $P < 0.0001$ ,  $r^2 = 0.78$ ;  $n = 24$ ), because this will stimulate greater production of proteases; third, addition of the synthetic signal to the wild type resulted in a significant increase in the expression of the *lasB* elastase structural gene, confirming an increased production of the elastase public goods that break down proteins ( $F_{(1,18)} = 302.2$ ,

$P < 0.0001$ ,  $n = 20$ ); fourth, addition of signal to the signal-blind strain resulted in no improvement in growth (Fig. 1a;  $F_{(1,30)} = 2.3$ ,  $P > 0.1$ ), as would be expected because the cells do not respond to the signal; and fifth, addition of purified elastase (porcine elastase) to the signal-blind strain increased growth (Fig. 1c;  $F_{(1,8)} = 116,800$ ,  $P < 0.0001$ ,  $n = 10$ ), by directly providing the public goods that allow proteins to be broken down.

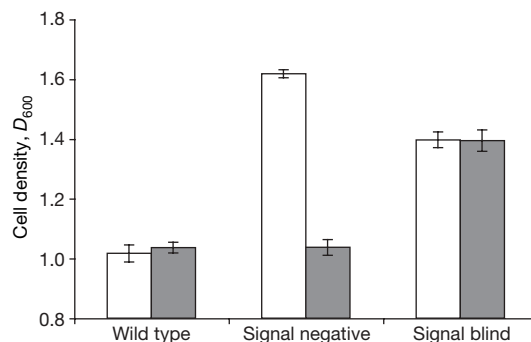
We also found that the production of the QS signal molecules and QS-dependent exoproducts (public goods) are metabolically costly. We compared the growth yield of the mutants and the wild type in nutrient-rich Luria–Bertani (LB) broth, in which the exoproducts produced by QS are not needed for growth. In these conditions, 22 h after inoculation, the QS mutants had grown to a higher density than the wild type (Fig. 2;  $F_{(1,34)} = 257.1$ ,  $P < 0.0001$ ). The cost of the response to QS molecules is further shown by the fact that addition of synthetic signal to the signal-negative mutant resulted in a decreased growth yield (Fig. 2;  $F_{(1,32)} = 1,043.32$ ,  $P < 0.0001$ ); it was significantly lower than that obtained with the signal-blind strain with or without signal (Fig. 2;  $F_{(1,33)} = 29.33$ ,  $P < 0.01$ ). Regulation of public goods production by QS generally occurs in the transition from exponential to stationary phase, and one would therefore expect fitness costs to occur primarily in the stationary phase. To test this, we examined growth of the wild type and QS mutants over time and found that growth rates were similar during exponential growth and that fitness differences occurred in the stationary phase (see Supplementary Fig. 3).

We then tested whether QS cheats can spread in a population, because they benefit from the signal or exoproducts produced by others without paying the production costs themselves<sup>1,2,11,14</sup>. We initiated populations of the wild type with a small proportion (1–3%) of one of the two mutants and followed their growth in QSM over 48 h. During this time, the signal-blind cheats increased in frequency from 1% to 45% and the signal-negative mutant increased from 3% to 66% (Fig. 3a). Consequently, over 48 h of growth, in mixed cultures, both the signal-blind and the signal-negative mutants had a significantly higher fitness than the parental wild type (in both cases,  $P = 0.02$ ,  $n = 6$ ; the relative fitness of the two mutants were 81-fold and 63-fold that of the wild type, respectively). Social evolution theory also predicts that the fitness of cheats in microbial populations should be frequency dependent, with relative fitness decreasing as cheats become more common<sup>22,23</sup>. As cooperators become more common, there is greater population growth and hence more opportunity for cheats to exploit cooperators<sup>22</sup>. In support of this prediction, we found that the relative fitness of the signal-blind cheat decreased significantly when it was more common (Fig. 3b;  $F_{(1,28)} = 57.15$ ,  $P < 0.0001$ ,  $n = 30$ ,  $r^2 = 0.671$ ).

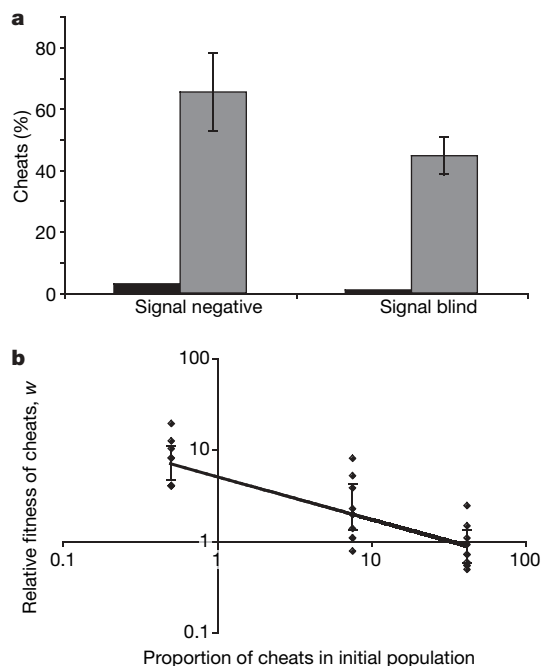
Our results show that QS is a social trait, susceptible to exploitation and invasion by cheats. Given this, how is QS maintained in



**Figure 1 | QS-controlled public goods provide a benefit at the population level in QSM.** **a**, The wild type grows (attenuance ( $D_{600}$ ) measurements after 24 h of growth) significantly more rapidly than the signal-blind and signal-negative mutants. Addition of synthetic signal improves the growth of the signal-negative mutant but not the signal-blind mutant (seven independent replicates per treatment). **b**, Increasing synthetic signal concentration improves cooperation and growth in a signal-negative mutant. The graph shows  $D_{600}$  measurements after 30 h of growth; six replicates per treatment. **c**, Addition of synthetic elastase significantly increases the growth of the signal-blind mutant after 20 h in QSM. All results are shown as means  $\pm$  s.e.m, five replicates per treatment.



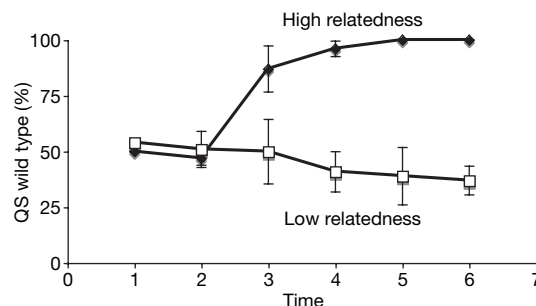
**Figure 2 | QS-dependent cooperation is costly.** Final cell density of wild type and mutants in a rich (LB) growth medium in the absence (open columns) and presence (filled columns) of signal (5  $\mu$ M 3O-C12-HSL). All results are shown as means  $\pm$  s.e.m. for six independent measurements per treatment.



**Figure 3 | QS is subject to cheating.** **a**, QS signal-negative and signal-blind cheats invade populations of wild-type cooperators over 48 h of growth in QSM. All results are shown as means  $\pm$  s.e.m. (estimated assuming binomial errors and logit transformation) for six independent measurements per treatment. Black and grey bars represent the starting and final percentage of cheats in the population, respectively. **b**, Invasion and relative fitness of signal-blind cheats are frequency dependent over 48 h of growth in QSM, with cheats having a higher fitness when they are less common. Relative fitness is the estimated growth rate of cheats relative to that of cooperators. The curve is a power regression line fitted to the raw data ( $n = 30$ ). All results are shown as means  $\pm$  s.e.m.

natural populations? The most likely explanation is kin selection<sup>8</sup>—if neighbouring cells tend to be close relatives they will have a shared interest in communicating honestly, and cooperating<sup>2,14</sup>. Kin selection is likely to be highly important in microbial social behaviours such as QS because of clonal reproduction and relatively local interactions<sup>11</sup>. We tested the importance of kin selection for QS with an experimental evolution approach. We maintained populations under conditions of relatively high and relatively low relatedness<sup>24</sup>, and examined how this influenced the relative fitnesses of the wild type and the signal-blind *lasR* mutant (see Supplementary Fig. 2). As predicted by kin selection theory<sup>14</sup>, QS was favoured at a relatively high relatedness (Fig. 4;  $F_{(1,4)} = 553.6$ ,  $P < 0.0001$ ,  $n = 6$ ). In the high-relatedness treatment, the wild type and the signal-blind mutant tended to be in different subpopulations, preventing exploitation of the wild type by cheats. In contrast, in the relatively low-relatedness treatment, subpopulations could contain both QS cooperators and the signal-blind cheats, allowing the signal-blind cheats to exploit the production of public goods (proteases) by the cooperative QS individuals and hence to increase in frequency within the population.

Our results and social evolution theory can help explain the prevalence of QS mutants in natural populations. First, in both environmental and clinical isolates, QS cheats are often found that do not perform normal QS behaviours<sup>20,25,26</sup>. The vast majority of these are signal-blind mutants that do not respond to signal. This is to be expected because signal-blind mutants avoid the cost of producing both the signal and exoproducts regulated by QS, whereas signal-negative mutants avoid only the cost of producing signal, which is likely to be far less metabolically costly than producing the exoproducts. Second, it has frequently been observed that such QS mutants are able to invade and spread in clinical settings, such as the lungs of



**Figure 4 | Kin selection and QS.** The proportion of quorum-sensing individuals (wild type) is plotted against time (rounds of growth). QS is favoured by higher relatedness: open squares, low relatedness; filled diamonds, high relatedness. All results are shown as means  $\pm$  s.e.m. for three independent replicate experiments per treatment.

humans with cystic fibrosis<sup>25,27</sup>. Populations of *P. aeruginosa* in the human lung are subject to low levels of dispersal in and out of the lung<sup>28</sup>, leading to relatively local competition within the lung. Such local competition reduces kin selection for cooperative behaviours such as QS, even if there are small numbers of lineages within the lung<sup>29</sup>. Consequently, kin selection theory predicts the observed pattern that QS cheats should invade over time in lung infections. More generally, we would predict a higher prevalence of QS cheats in any situation in which the average relatedness between interacting cells was lower or in which there was a greater intensity of local competition.

## METHODS SUMMARY

**Strains and growth conditions.** The *P. aeruginosa* strains PAO1 *lecA::lux* or PAO1 were used as QS-positive cooperating strains; *lasI* (signal-negative) and *lasR* (signal-blind) mutants were used as QS cheats. Strains were routinely grown in LB broth. For growth competition assays we used QSM. This consisted of OS minimal medium<sup>30</sup> supplemented with 1% BSA (Sigma) as the sole carbon source. For growth-curve experiments we grew strains in 300  $\mu$ l of LB broth or QSM in a combined automated luminometer–spectrometer (GENios Pro; Tecan Group). The QS experimental evolution study to test whether selection for QS could be influenced by indirect fitness benefits was based on the global competition treatments of a previous experiment<sup>24</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 29 June; accepted 18 September 2007.

- Keller, L. & Surette, M. G. Communication in bacteria: an ecological and evolutionary perspective. *Nat. Rev. Microbiol.* **4**, 249–258 (2006).
- Diggle, S. P., Gardner, A., West, S. A. & Griffin, A. S. Evolutionary theory of bacterial quorum sensing: when is a signal not a signal? *Phil. Trans. R. Soc. B* **362**, 1241–1249 (2007).
- Henke, J. M. & Bassler, B. L. Bacterial social engagements. *Trends Cell Biol.* **14**, 648–656 (2004).
- Williams, P., Winzer, K., Chan, W. C. & Camara, M. Look who's talking: communication and quorum sensing in the bacterial world. *Phil. Trans. R. Soc. B* **362**, 1119–1134 (2007).
- Shapiro, J. A. Thinking about bacterial populations as multicellular organisms. *Annu. Rev. Microbiol.* **52**, 81–104 (1998).
- Maynard Smith, J. & Harper, D. *Animal Signals* (Oxford Univ. Press, Oxford, 2003).
- Lehmann, L. & Keller, L. The evolution of cooperation and altruism—a general framework and a classification of models. *J. Evol. Biol.* **19**, 1365–1376 (2006).
- Hamilton, W. D. Genetical evolution of social behaviour. I & II. *J. Theor. Biol.* **7**, 1–52 (1964).
- Frank, S. A. *Foundations of Social Evolution* (Princeton Univ. Press, Princeton, NJ, 1998).
- Ratnieks, F. L. W., Foster, K. R. & Wenseleers, T. Conflict resolution in insect societies. *Annu. Rev. Entomol.* **51**, 581–608 (2006).
- West, S. A., Griffin, A. S., Gardner, A. & Diggle, S. P. Social evolution theory for microorganisms. *Nat. Rev. Microbiol.* **4**, 597–607 (2006).
- West, S. A., Griffin, A. S. & Gardner, A. Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *J. Evol. Biol.* **20**, 415–432 (2007).

13. Redfield, R. J. Is quorum sensing a side effect of diffusion sensing? *Trends Microbiol.* **10**, 365–370 (2002).
14. Brown, S. P. & Johnstone, R. A. Cooperation in the dark: signalling and collective action in quorum-sensing bacteria. *Proc. R. Soc. Lond. B* **268**, 961–965 (2001).
15. Venturi, V. Regulation of quorum sensing in *Pseudomonas*. *FEMS Microbiol. Rev.* **30**, 274–291 (2006).
16. Schuster, M., Lostroh, C. P., Ogi, T. & Greenberg, E. P. Identification, timing, and signal specificity of *Pseudomonas aeruginosa* quorum-controlled genes: a transcriptome analysis. *J. Bacteriol.* **185**, 2066–2079 (2003).
17. Pesci, E. C., Pearson, J. P., Seed, P. C. & Iglewski, B. H. Regulation of *las* and *rhl* quorum sensing in *Pseudomonas aeruginosa*. *J. Bacteriol.* **179**, 3127–3132 (1997).
18. Latifi, A. *et al.* A hierarchical quorum-sensing cascade in *Pseudomonas aeruginosa* links the transcriptional activators LasR and RhlR (VsmR) to expression of the stationary-phase sigma factor RpoS. *Mol. Microbiol.* **21**, 1137–1146 (1996).
19. Fletcher, M. P. *et al.* A dual biosensor for 2-alkyl-4-quinolone quorum sensing signal molecules. *Environ. Microbiol.* **9**, 2683–2693 (2007).
20. Schaber, J. A. *et al.* Analysis of quorum sensing-deficient clinical isolates of *Pseudomonas aeruginosa*. *J. Med. Microbiol.* **53**, 841–853 (2004).
21. Chhabra, S. R. *et al.* Synthetic analogues of the bacterial signal (quorum sensing) molecule *N*-(3-oxododecanoyl)-L-homoserine lactone as immune modulators. *J. Med. Chem.* **46**, 97–104 (2003).
22. Ross-Gillespie, A., Gardner, A., West, S. A. & Griffin, A. S. Frequency dependence and cooperation: theory and a test with bacteria. *Am. Nat.* **170**, 331–342 (2007).
23. MacLean, R. C. & Gudelj, I. Resource competition and social conflict in experimental populations of yeast. *Nature* **441**, 498–501 (2006).
24. Griffin, A. S., West, S. A. & Buckling, A. Cooperation and competition in pathogenic bacteria. *Nature* **430**, 1024–1027 (2004).
25. Smith, E. E. *et al.* Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. *Proc. Natl Acad. Sci. USA* **103**, 8487–8492 (2006).
26. Denervaud, V. *et al.* Characterization of cell-to-cell signaling-deficient *Pseudomonas aeruginosa* strains colonizing intubated patients. *J. Clin. Microbiol.* **42**, 554–562 (2004).
27. Salunkhe, P. *et al.* A cystic fibrosis epidemic strain of *Pseudomonas aeruginosa* displays enhanced virulence and antimicrobial resistance. *J. Bacteriol.* **187**, 4908–4920 (2005).
28. Jones, A. M. *et al.* Identification of airborne dissemination of epidemic multiresistant strains of *Pseudomonas aeruginosa* at a CF centre during a cross infection outbreak. *Thorax* **58**, 525–527 (2003).
29. West, S. A., Pen, I. & Griffin, A. S. Conflict and cooperation—cooperation and competition between relatives. *Science* **296**, 72–75 (2002).
30. Heurlier, K. *et al.* Quorum-sensing-negative (*lasR*) mutants of *Pseudomonas aeruginosa* avoid cell lysis and death. *J. Bacteriol.* **187**, 4875–4883 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank P. Williams and R. Chhabra for the provision of AHL compounds; D. Jackson for comments on the manuscript; and A. Ross-Gillespie for help in analysing the data. We acknowledge funding from the Royal Society and the Leverhulme Trust.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to S.P.D. ([steve.diggle@nottingham.ac.uk](mailto:steve.diggle@nottingham.ac.uk)) or S.A.W. ([stu.west@ed.ac.uk](mailto:stu.west@ed.ac.uk)).



## METHODS

**Bacterial strains and growth conditions.** We grew bacteria in LB broth or on *Pseudomonas* isolation agar (Difco). The *P. aeruginosa* strains PAO1 *lecA::lux<sup>+</sup>* or PAO1 were used as QS-positive cooperating strains. The construction of *lasI* (signal-negative) and *lasR* (signal-blind) cheating mutants in PAO1 *lecA::lux* is described elsewhere<sup>19</sup>. Although the *lasI* mutant (PAO1 *lecA::luxΔlasI*) is unable to produce the 3O-C12-HSL signal<sup>19</sup>, it still responds to the presence of signal, as shown by the fact that in the presence of the 3O-C12-HSL signal (20 μM), but not in its absence, it significantly expresses the QS-dependent *lecA* gene (which codes for a galactophilic lectin)<sup>31</sup> ( $n = 12$ ,  $P < 0.0001$ ) and produces the QS-dependent protease elastase ( $n = 6$ ,  $P < 0.0001$ ). The *lasR* mutant (PAO1 *lecA::luxΔlasR*) produces minimal amounts of signal<sup>19</sup> but does not respond to extracellular signal as shown by the fact that *lecA* expression ( $n = 12$ ,  $P = 0.919$ ) and elastase production ( $n = 6$ ,  $P = 0.072$ ) are not restored on the provision of signal.

We routinely grew all strains at 37 °C in 5 ml of LB broth in 20-ml tubes (Sterilin) with shaking at 200 r.p.m. QSM was made by using OS minimal medium<sup>30</sup> supplemented with 1% BSA (Sigma) as the sole carbon source. For growth curve experiments, we grew strains in 300 μl of LB broth or QSM in a combined automated luminometer–spectrometer (GENios Pro; Tecan Group). Growth was followed by the determination of  $D_{600}$  values, and viable counts (colony-forming units) were obtained by plating suitable dilutions made in saline (0.9% NaCl) onto LB agar. Where required, we added synthetic AHL signal<sup>21,32</sup> at the concentrations indicated.

Unless stated otherwise, we performed all analyses by model simplification to the minimum adequate model, using generalized linear modelling techniques implemented in GLMStat 6.0 (Kagi Shareware). When analysing the effect of adding signal on the signal-negative mutant, we used  $\log_e(\text{signal} + 1)$  logging to linearize values and then added 1 to account for zeros.

**Growth competition assays.** We determined that protease production is an appropriate public-goods cooperative trait for our experiments. We measured the growth rates of a wild-type (cooperator) strain that produces QS-dependent proteases and two mutant strains (cheats) that either do not make QS signal but can respond to signal and thus make proteases (signal-negative) or do make minimal amounts of signal and do not respond (signal-blind). For growth competition assays, we first grew PAO1 (QS positive cooperator) and PAO1 *lecA::luxΔlasI* and PAO1 *lecA::luxΔlasR* (cheats) overnight separately in LB broth. Attenuance ( $D_{600}$ ) readings were performed, and PAO1 and the relevant cheat were then mixed together in the desired ratio (namely 99:1). To determine the exact starting ratio, a sample was then immediately taken and we performed serial dilutions to obtain single colonies on LB agar plates. These were incubated overnight at 37 °C and the ratio of cheats was determined with a light camera (E. G. & G. Berthold): cheats contained the *lux* genes and were therefore bioluminescent, whereas PAO1 was not. From the initial inoculum we transferred 20 μl to 200 μl of QSM and this was incubated for 48 h at 37 °C. Samples were then taken, serial dilutions were performed to obtain single colonies, and the cheat ratio was determined as above.

These data were analysed with a non-parametric sign test, because relative fitness ( $w$ ) is non-normally distributed (Shapiro–Wilkinson test; signal-negative,  $P = 0.001$ ; signal-blind,  $P = 0.02$ ; Fig. 3a). In this experiment, the fitness of the signal-blind mutant was about 81-fold that of the wild type, and the fitness of the signal-negative mutant was about 63-fold that of the wild type. Furthermore, our estimates of the relative fitness of the QS mutants are underestimates because they must also pay the metabolic cost of expression of the *lux* reporter. We tested for frequency dependence by examining how the relative fitness of the signal-blind mutant changed as its initial starting frequency in the population was varied from about 1% to 50% (Fig. 3b). For frequency-dependence experiments, ratios of about 1%, 10% and 50% cheats (signal-blind) were mixed with PAO1. From these inocula we transferred 20 μl to 200 μl of QSM and incubated for 48 h at 37 °C. Samples were then taken, serial dilutions were performed to obtain single colonies, and the cheat ratio was determined as above.

**Selection experiment.** We performed an experimental evolution study to test whether selection for QS can be influenced by kin selection. Our design is based

on the global competition treatments of a previous experiment<sup>24</sup> (see Supplementary Fig. 2). We used the signal-blind *lasR* mutant because this is the mutant most commonly observed in nature<sup>20,25–27</sup>. Our experiment involved two treatments: relatively high relatedness and relatively low relatedness<sup>24</sup>. Each treatment was replicated three times. Within each replicate we split each population into 12 subpopulations. We initiated each subpopulation with a 1:1 mix of the wild type and the signal-blind mutant. We did this by growing PAO1 wild type (QS positive cooperator) and the signal-blind strain (cheat) individually overnight in LB medium, taking attenuation measurements to measure cell density, and then mixing the strains in a single tube at a 1:1 ratio. We then added 20-μl aliquots of the 1:1 mix into 12 subpopulations in wells of a 96-well plate, each containing 200 μl of QSM, and incubated this plate for 48 h at 37 °C. We left these subpopulations to grow in QSM for 48 h, after which the populations were pooled and used to initiate 12 new subpopulations, and the process was repeated. We measured the ratio of cooperators to cheats in the mixed population by serially diluting the mixed tube, plating on LB agar plates to obtain single colonies, incubating these plates for 24 h, and viewing with a light camera. We manipulated relatedness by varying how the mix of the subpopulations was used to initiate the 12 new subpopulations. We achieved relatively high relatedness by initiating each subpopulation with a single colony, taken at random from a sample of the subpopulation mix. Specifically, we inoculated single colonies into 12 wells containing LB medium, incubated these at 37 °C for 24 h and then used 20 μl of culture from each well to inoculate a new subpopulation. For low relatedness we conducted the same procedure, except that the initial inoculation was performed with 20 μl of the pooled sample (about  $1.2 \times 10^9$  cells), rather than a single colony. We then incubated these subpopulations for 48 h at 37 °C, as described above, and repeated this procedure for a total of six rounds of growth in subpopulations. Before data analysis, we arcsin square root transformed the proportion of cooperators to account for the non-normality of proportion data.

**Time and cell-density-dependent measurement of bioluminescence.** We determined bioluminescence as a function of cell population density with a combined automated luminometer–spectrometer (GENios Pro; Tecan Group). We diluted overnight cultures of *P. aeruginosa* 1:1,000 (in LB medium) or 1:20 (in QSM), and 0.2-ml cultures were grown in 96-well microtitre plates. Luminescence and turbidity were automatically determined every 30 min. Luminescence was measured in relative light units. Where required, we supplied AHL signal at the concentrations indicated. AHLs were synthesized as described previously<sup>21</sup>. AHL was dissolved in methanol before it was added to growth medium at the indicated concentrations.

**Calculation of relative fitness ( $w$ ) of strains.** We calculated the relative fitness of cheating mutants ( $w$ ) by comparing the frequency of cheats at the beginning and end of the experiment. Specifically,  $w$  is given by  $[x_2(1 - x_1)]/[x_1(1 - x_2)]$ , where  $x_1$  is the initial proportion of cheats in the population and  $x_2$  is their final proportion<sup>22</sup>. For example,  $w = 2$  would correspond to the mutant's growing twice as fast as the wild-type cooperator.

**Assay for elastolytic activity.** We determined the elastolytic activity of bacterial supernatants by using the elastin Congo red (ECR; Sigma) assay as described previously<sup>33</sup>. We added a 100-μl aliquot of bacterial supernatant to 900 μl of ECR buffer (100 mM Tris-HCl pH 7.5, 1 mM CaCl<sub>2</sub>) containing 20 mg of ECR and incubated the mixture with shaking at 37 °C for 3 h. We removed insoluble ECR by centrifugation, and the  $A_{495}$  of the supernatant was measured. LB medium was used as a negative control.

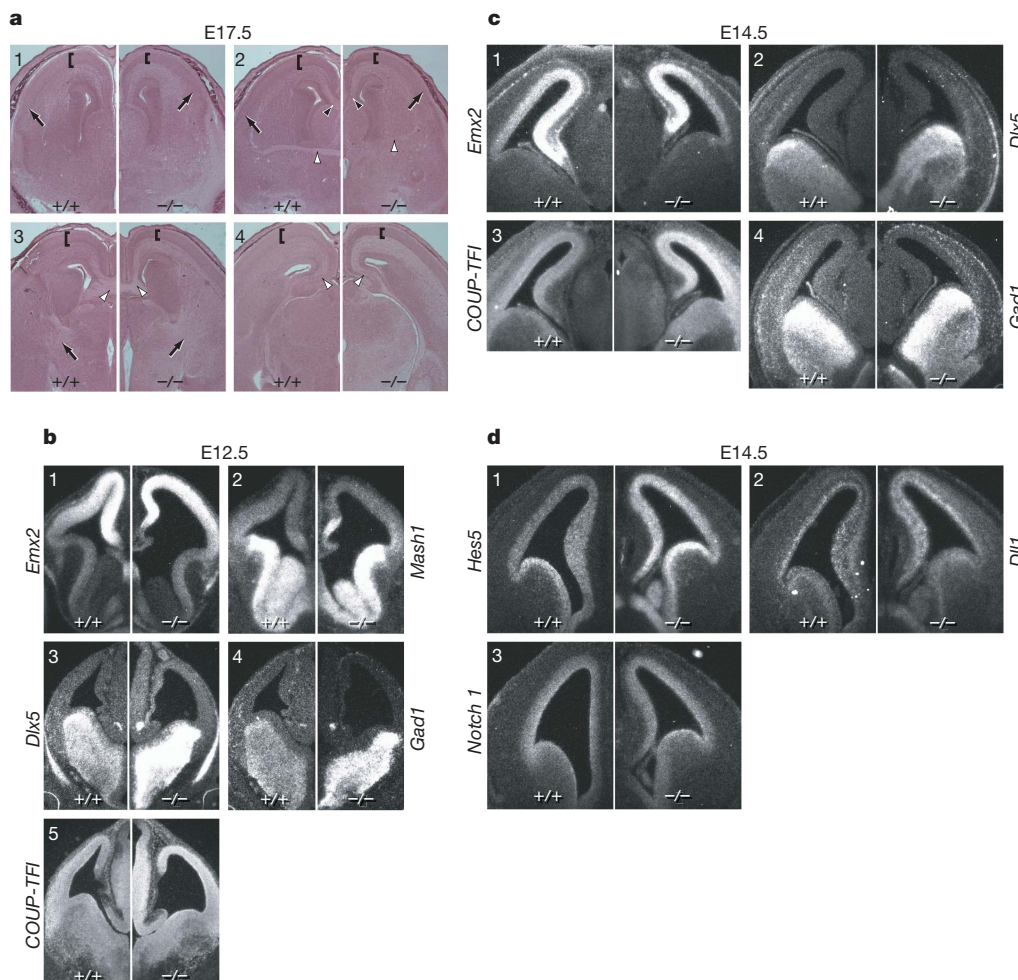
31. Winzer, K. *et al.* The *Pseudomonas aeruginosa* lectins PA-IL and PA-III are controlled by quorum sensing and by RpoS. *J. Bacteriol.* **182**, 6401–6411 (2000).
32. Chhabra, S. R. *et al.* Autoregulation of carbapenem biosynthesis in *Erwinia carotovora* by analogs of *N*-(3-oxohexanoyl)-L-homoserine lactone. *J. Antibiot. (Tokyo)* **46**, 441–454 (1993).
33. Diggle, S. P. *et al.* Advancing the quorum in *Pseudomonas aeruginosa*: MvaT and the regulation of *N*-acylhomoserine lactone production and virulence gene expression. *J. Bacteriol.* **184**, 2576–2586 (2002).

# SMRT-mediated repression of an H3K27 demethylase in progression from neural stem cell to neuron

Kristen Jepsen<sup>1†</sup>, Derek Solum<sup>1†</sup>, Tianyuan Zhou<sup>1†</sup>, Robert J. McEvilly<sup>1</sup>, Hyun-Jung Kim<sup>1</sup>, Christopher K. Glass<sup>2</sup>, Ola Hermanson<sup>3</sup> & Michael G. Rosenfeld<sup>1</sup>

A series of transcription factors critical for maintenance of the neural stem cell state have been identified<sup>1–3</sup>, but the role of functionally important corepressors<sup>4–7</sup> in maintenance of the neural stem cell state and early neurogenesis remains unclear. Previous studies have characterized the expression of both *SMRT* (also known as *NCoR2*, nuclear receptor co-repressor 2) and *NCoR* in a variety of developmental systems<sup>8</sup>; however, the specific role of the *SMRT* corepressor in neurogenesis is still to be determined. Here we report a critical role for *SMRT* in forebrain development and in maintenance of the neural stem cell state. Analysis of a series of

markers in *SMRT*-gene-deleted mice revealed the functional requirement of *SMRT* in the actions of both retinoic-acid-dependent and Notch-dependent forebrain development. In isolated cortical progenitor cells, *SMRT* was critical for preventing retinoic-acid-receptor-dependent induction of differentiation along a neuronal pathway in the absence of any ligand. Our data reveal that *SMRT* represses expression of the jumonji-domain containing gene *JMJD3*, a direct retinoic-acid-receptor target that functions as a histone H3 trimethyl K27 demethylase and which is capable of activating specific components of the neurogenic program.



**Figure 1 | Impaired neural development in *SMRT*<sup>-/-</sup> embryos.** **a**, Matched haematoxylin and eosin (H&E)-stained cryosections from E17.5 wild-type or *SMRT*<sup>-/-</sup> cortex. Brackets indicate cortical plate in panels 1 to 4. Black arrowheads indicate corpus callosum (panel 2). White arrowheads indicate anterior commissure (panel 2), hippocampal commissure (panel 3), and hippocampal plate (panel 4). Arrows indicate neocortex/palaeocortex boundary (panels 1, 2) and corticothalamic and thalamocortical projections (panel 3). **b**, *In situ* analysis of wild-type and *SMRT*<sup>-/-</sup> E12.5 embryos using the indicated radiolabelled anti-sense RNA probes. **c**, **d**, *In situ* analysis of wild-type and *SMRT*<sup>-/-</sup> E14.5 embryos using the indicated radiolabelled anti-sense RNA probes.

<sup>1</sup>Howard Hughes Medical Institute, Department of Medicine, <sup>2</sup>Department of Cellular and Molecular Medicine, University of California, San Diego, School of Medicine, 9500 Gilman Drive, La Jolla, California 92093, USA. <sup>3</sup>Center of Excellence in Developmental Biology (CEDB/DBRM), Organic Bioelectronics (OBOE), Department of Neuroscience, Karolinska Institutet, SE17177 Stockholm, Sweden.

†These authors contributed equally to this work.



To investigate roles of SMRT in neurogenesis, we generated mice gene-deleted for *SMRT* (Supplementary Fig. 1a–d). Although *SMRT*<sup>−/−</sup> mice generally died before embryonic day (E)16.5 owing to a lethal heart defect, myocyte-specific targeting of *SMRT* allowed rescue of the heart defect and survival to birth (Supplementary Fig. 1e). Inspection of forebrain structures at E17.5 revealed significant differences between wild-type and *SMRT*<sup>−/−</sup> cortex, most noticeably with a shift in the proportional volumes of dorsal and ventral telencephalon regions. A thinning of the neocortical plate was apparent throughout both the medial and lateral regions, and extended throughout the rostro-caudal axis of the brain (Fig. 1a, and Supplementary Fig. 2), extending medially to include the hippocampus (Fig. 1a, and panel 4 of Supplementary Fig. 2). The ventro-lateral boundaries of the neocortex and the palaeocortex terminated in more dorsal positions in *SMRT*<sup>−/−</sup> brain (Fig. 1a, and panels 1, 2 of Supplementary Fig. 2). Axonal projections linking opposite cortical hemispheres were also significantly altered (Fig. 1a, and panels 2–4 of Supplementary Fig. 2), with both the anterior commissure and the corpus callosum exhibiting near-complete loss of volume and the hippocampal commissure also significantly reduced. Interestingly, axonal projections between the thalamus and cortex appeared more highly represented in the *SMRT*<sup>−/−</sup> brain (Fig. 1a, and panel 3 of Supplementary Fig. 2).

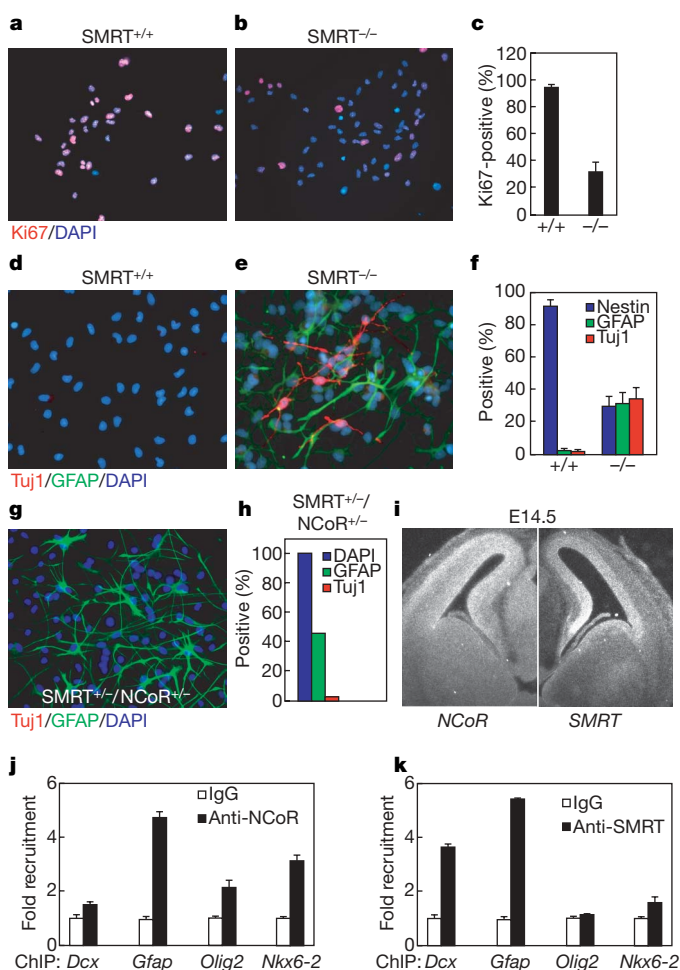
At E12.5, levels of *Emx2* and *Mash1* (panels 1, 2 of Fig. 1b), which are respectively representative of the dorsal and ventral telencephalon<sup>1,9,10</sup>, were unchanged in *SMRT*<sup>−/−</sup> brain, suggesting that the initial specification of dorsal and ventral regions occurred normally in the absence of SMRT, and results obtained with additional dorsal markers *Emx1*, *Ngn2* and *Pax6* and ventral markers *Dlx2* and *Gsh2* (data not shown)<sup>1,9,10</sup> further supported this conclusion. Analysis of regional subdivisions of the ventral telencephalon with markers specific for medial ganglionic eminence (*Lhx6*, *Lhx7*, *Nkx2.1*)<sup>1,9,10</sup> and the lateral and caudal ganglionic eminence (*Oct6*, *Isl1*)<sup>1,9,10</sup> also revealed expected distribution patterns within these structures (data not shown). Interestingly, markedly increased levels of expression of *Dlx5* and *Gad1*, markers of differentiated interneurons<sup>11</sup>, were observed in the ventral telencephalon of E12.5 *SMRT*<sup>−/−</sup> brain, together with higher levels of *COUP-TFI*, an early marker of differentiating neurons, in dorsal telencephalon and thalamus (panels 3–5 of Fig. 1b).

A similar panel of analyses at E14.5 again revealed the expected distribution of markers delineating dorsal and ventral boundaries and specific subregions (Fig. 1c and data not shown), although by this stage the hippocampus, neocortex, and palaeocortex of the *SMRT*<sup>−/−</sup> brain were all significantly smaller than equivalent regions in wild-type littermates, as illustrated by *EMX2* and *COUP-TFI* expression (panels 1 and 3 of Fig. 1c). In the *SMRT*<sup>−/−</sup> brain at this stage, the expression of *COUP-TFI* was substantially upregulated in dorsal telencephalon, as well as more broadly distributed both dorsally and ventrally, in the mediolateral cortex and caudal ganglionic eminence, respectively, whereas the expression levels of *Dlx5* and *Gad1* in ventral telencephalon had resolved to levels similar to wild-type (panels 2, 4 of Fig. 1c).

A variety of mechanisms have been genetically implicated as influencing patterning and differentiation in telencephalon development, including retinoic acid (RA) signalling<sup>12</sup>. RA at early stages influences the initial expression of ligands involved in the Wnt/BMP and SHH/FGF signalling pathways, which subsequently regulate dorsal and ventral telencephalic development<sup>12</sup>, and, at later stages it is implicated in refinement of the dorsoventral telencephalic boundary<sup>13</sup>. Given that *COUP-TFI*, itself a target of RA signalling during brain development<sup>14</sup>, was upregulated in the *SMRT*<sup>−/−</sup> telencephalon (panel 5 of Fig. 1b, and panel 3 of Fig. 1c), and that SMRT has been suggested preferentially to mediate repression of unliganded RA receptors<sup>15</sup>, we examined potential effects of enhanced RA signalling on the signalling pathways affecting early brain development, as well as on Notch signalling, where SMRT and NCoR have been shown to repress Notch targets through interaction with the DNA binding protein CSL (CBF-1, suppressor of hairless, and Lag-1)<sup>7,16–18</sup>. While

the SHH/FGF (*SHH*, *Gli1*, *Gli3*, *Nkx2.1*) and Wnt (*Wnt3*, *Wnt7A*, *FOXG1*) pathways were unaffected (data not shown), we did record significantly elevated levels of expression for *Hes5* and *Dll1*, but not for *Notch1*, at E14.5 in the ventricular zones of *SMRT*<sup>−/−</sup> telencephalon (Fig. 1d). These data are consistent with chromatin immunoprecipitation (ChIP) experiments performed in embryonic telencephalon that reveal the presence of SMRT on both *Hes5* and *Dll1* promoters, suggesting direct transcriptional repression of these Notch target genes by SMRT (Supplementary Fig. 3a).

We next investigated the role of SMRT in maintenance of the neural stem cell phenotype, employing cortical progenitor cells from individual E13 mouse embryos that were isolated and cultured as previously described<sup>7</sup>. In the presence of FGF2, progenitor cells from wild-type embryos proliferated as assessed by the expression of the proliferative marker Ki67, and formed colonies of self-renewing, nestin-positive neural stem cells (Fig. 2a, c, f). While cortical progenitors derived from E13 *SMRT*<sup>−/−</sup> embryos were initially indistinguishable from wild-type cells, after four to six days in culture, *SMRT*<sup>−/−</sup>



**Figure 2 | SMRT is a regulator of neural stem cell state.**

**a–c**, Immunofluorescent labelling and quantification ( $\pm$  standard error of the mean, s.e.m.) of Ki67-expressing cells (red) in *SMRT*<sup>+/+</sup> and *SMRT*<sup>−/−</sup> neural stem cells after three days in culture. Nuclear 4,6-diamidino-2-phenylindole (DAPI) staining in blue. **d, e**, Immunofluorescent labelling of GFAP-expressing (green) and Tuj1<sup>+</sup> (red) cells in *SMRT*<sup>+/+</sup> and *SMRT*<sup>−/−</sup> neural stem cells after five days in culture. **f**, Quantification of nestin-expressing, GFAP-expressing and Tuj1-positive cells in *SMRT*<sup>+/+</sup> and *SMRT*<sup>−/−</sup> neural stem cells ( $\pm$  s.e.m.). **g, h**, Immunofluorescent labelling and quantification of GFAP-expressing (green) neural stem cells from *SMRT*<sup>+/+</sup>/NCoR<sup>+/−</sup> embryos. **i**, *In situ* analysis in wild-type E14.5 cortex using RNA antisense probes to NCoR or SMRT. **j, k**, ChIP in neural stem cells using anti-SMRT or anti-NCoR antibodies on neuronal (*Dcx*), astroglial (*Gfap*) or oligodendrocyte (*Olig2*, *Nkx6-2*) promoters ( $\pm$  standard deviation, s.d.).



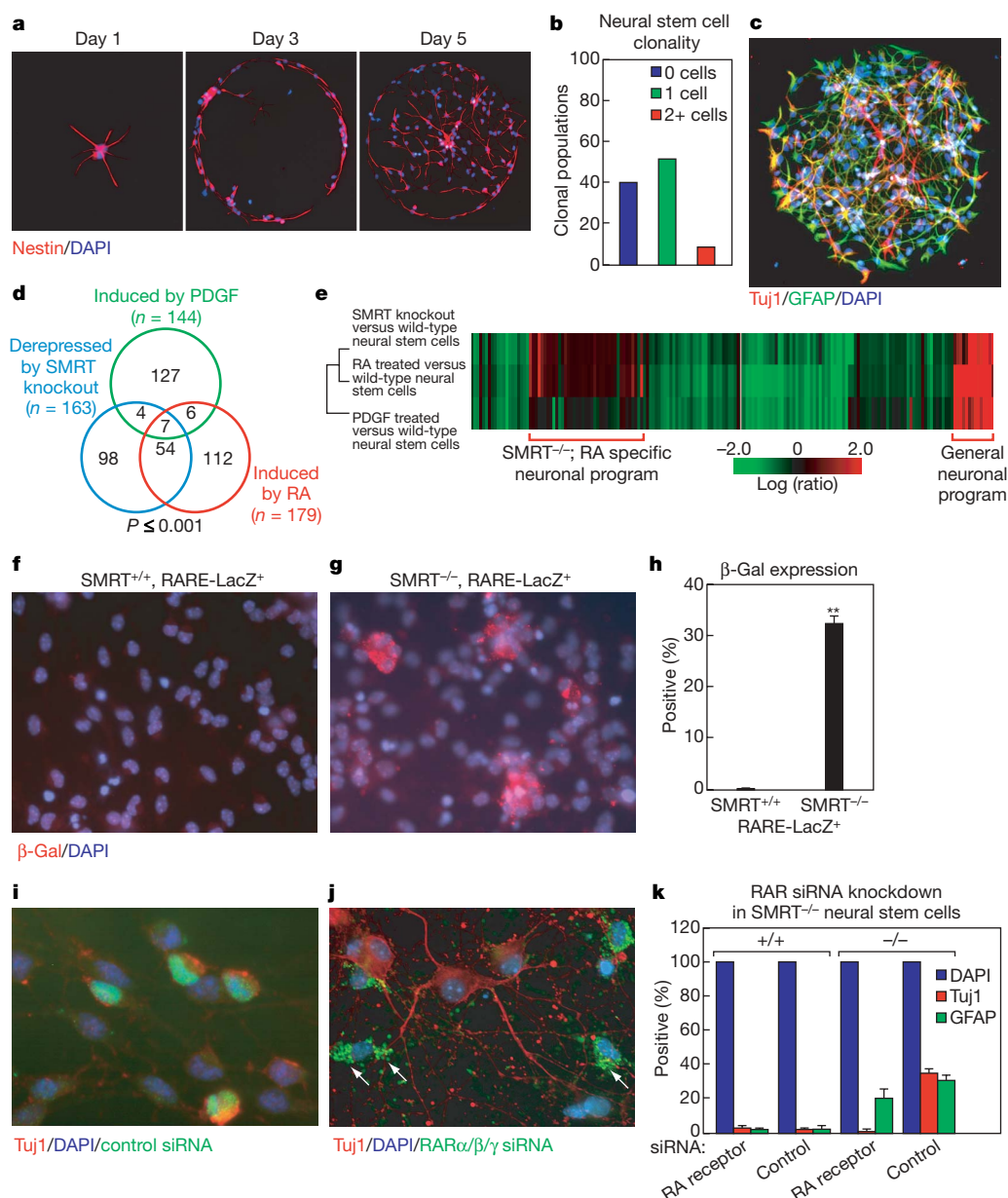
cells exhibited impaired proliferation (Fig. 2b, c) with no differences in apoptosis (Supplementary Fig. 3c, d). Furthermore, many *SMRT*<sup>-/-</sup>-derived progenitor cells displayed morphological signs of differentiation and expressed markers of either neuronal (Tuj1-positive) or astroglial (GFAP-positive) differentiation, with a corresponding decrease in the number of nestin-expressing cells (Fig. 2d–f).

Gene deletion of the SMRT-related corepressor, *NCoR*, results in cortical progenitors with impaired self-renewal that spontaneously differentiate into GFAP-expressing astroglia-like cells<sup>7</sup>. Interestingly, cortical progenitors from *SMRT*<sup>+/-</sup>/*NCoR*<sup>+/-</sup> double heterozygotes spontaneously differentiated into astroglia-like cells, as indicated by GFAP expression (Fig. 2g, h), while those from either single heterozygote were phenotypically indistinguishable from the wild type (Supplementary Fig. 3e, f). These data suggest that regulation of the astroglial pathway depends on proper gene-dosage of both *NCoR* and *SMRT*, while differentiation along the neuronal pathway appears to be regulated by loss of *SMRT* alone. We note that *SMRT* messenger RNA is primarily expressed in the ventricular zone region, both in the dorsal and ventral telencephalic regions, where multipotent neural precursors reside, whereas the expression of *NCoR* is more broadly distributed throughout the dorsal cortex with its signal most pronounced in the outer cortical plate and in a narrow band

that probably corresponds to neuroepithelium, and is expressed at much lower levels in the ventral telencephalon, including the ventricular zone (Fig. 2i). Additionally, ChIP analysis revealed that while *NCoR* and *SMRT* were equivalently recruited to the glial-specific *Gfap* promoter, *SMRT* was preferentially recruited to the promoter of the neuronal gene *Dcx* (Fig. 2j, k).

To determine whether the increased number of Tuj1- and GFAP-expressing cells was a result of differentiation of multipotent progenitors or was due to an increase in the number of neuronal- or glial-restricted progenitors in *SMRT*<sup>-/-</sup> cortical progenitors, microdots of poly-L-ornithine and fibronectin were sprayed on culture dishes to obtain discrete points of potential adhesion of neural stem cells, providing containment of growth, because progenitor cells will not grow on uncoated areas (Fig. 3a)<sup>7</sup>. After five days in culture, while less than 10% of the microdots originally contained more than one cell (Fig. 3b), all *SMRT*<sup>-/-</sup> microdot colonies contained cells expressing the three cell-type markers (nestin, Tuj1 and GFAP), indicating that single E13 cortical progenitor cells derived from *SMRT*<sup>-/-</sup> embryos were multipotent (Fig. 3c).

To identify SMRT-regulated genes involved in early neurogenic events, RNA profiling experiments were performed, comparing RNA derived from E13 *SMRT*<sup>-/-</sup> progenitor cell cultures to untreated



**Figure 3 | Individual *SMRT*<sup>-/-</sup> neural stem cells are multipotent and require RA receptor for neuronal differentiation.** **a**, Growth of nestin-positive (red) *SMRT*<sup>-/-</sup> neural stem cell colonies are restricted to poly-L-ornithine-coated and fibronectin-coated microdots. **b**, Percentage of microdots visualized by anti-fibronectin labelling containing 0, 1 or 2 cells upon initial plating. **c**, Representative colony from *SMRT*<sup>-/-</sup> neural stem cells demonstrating the presence of neuronal (red, Tuj1<sup>+</sup>) and astroglial (green, GFAP-expressing) cells in a single colony. **d**, **e**, RNA profiling data from *SMRT*<sup>-/-</sup> neural stem cells and PDGF- or RA-treated wild-type neural stem cell cultures with the number of overlapping upregulated genes indicated (10 h).

**f–h**, Immunofluorescent labelling and quantification of β-Gal expression (red) on *SMRT*<sup>+/+</sup> or *SMRT*<sup>-/-</sup> cells carrying an *RARE-lacZ* transgene (± s.e.m., \*\**p* < 0.005).

**i–k**, Immunofluorescent labelling and quantification of siRNA-transfected (green) and Tuj1<sup>+</sup> (red) neural stem cells following lipofection with either FITC-labelled control siRNA alone or in combination with siRNAs targeting *RARα*, *RARβ* and *RARγ* (± s.e.m.).

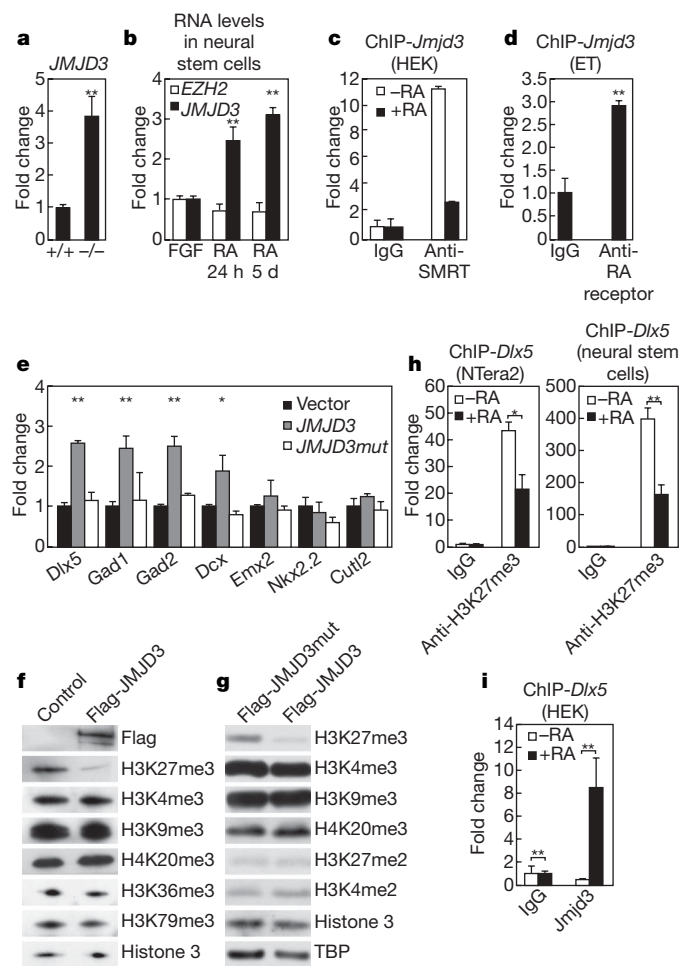
wild-type cells 10 h after plating, and to wild-type cells treated for 10 h with RA or platelet-derived growth factor (PDGF). A number of genes characteristic of neuronal identity as well as genes encoding a diverse set of growth factors implicated in neural cell growth and survival were upregulated in *SMRT*<sup>-/-</sup> cells (Supplementary Figs 4 and 5a). This gene activation profile is unique to the specific cell lineage milieu, because depletion of *SMRT* by siRNA in U2OS osteoblastoma cells did not entirely mimic what was observed in neural stem cells (Supplementary Fig. 5b, c). When compared with wild-type progenitor cell cultures treated with RA, we observed that approximately one-third of RA-induced genes were also induced in *SMRT*<sup>-/-</sup> cells, compared to only 5% of PDGF-induced genes (Fig. 3d, e). Consistent with RNA profiling data, we detected  $\beta$ -galactosidase ( $\beta$ -Gal) expression in approximately one-third of cortical progenitor cells derived from *SMRT*<sup>-/-</sup> embryos carrying an RA responsive element (*RARE-lacZ* transgene<sup>19</sup> in the absence of RA, while virtually no  $\beta$ -Gal expression was detected in cells derived from *SMRT*<sup>+/+</sup>, *RARE-lacZ* embryos in the absence of RA (Fig. 3f–h). These data suggest that ligand-independent RA-receptor activation function can occur in the absence of *SMRT*.

To establish that RA receptor is necessary for the spontaneous neuronal differentiation observed in *SMRT*<sup>-/-</sup> derived cultures, *SMRT*<sup>-/-</sup> cells were lipofected with a mix of specific siRNAs to *RAR $\alpha$* , *RAR $\beta$*  and *RAR $\gamma$*  (Supplementary Fig. S5d), in concert with fluorescein isothiocyanate (FITC)-labelled control small interfering (si)RNA to mark transfected cells. Indeed, when RA receptors are depleted from a subset of cells in *SMRT*<sup>-/-</sup> cultures, these cells were unable to differentiate into Tuj1-positive neuronal cells (Fig. 3i–k), but maintained the ability to differentiate into GFAP-positive astroglial cells (Fig. 3k). In contrast, in *SMRT*<sup>-/-</sup> cells transfected with control siRNA alone, approximately one-third of cells were Tuj1-positive, as was observed in untransfected cultures (Figs 3k and 2f). siRNA-positive cells appeared to lack Tuj1-positive processes; this seemed to be specifically associated with the uptake of siRNA, because Tuj1-positive processes were observed in siRNA-negative cells (Fig. 3i, j). To determine whether individual genes upregulated in *SMRT*<sup>-/-</sup> cortical progenitors were sufficient to drive the neurogenic program in the neural stem cells, we ectopically expressed individual complementary DNAs in the neural stem cells, thus mimicking the upregulation of genes observed in *SMRT*<sup>-/-</sup> cells (Fig. 4, and Supplementary Fig. 6). One such gene, *JMJD3*, a member of the jumonji C family of putative histone demethylases<sup>20–22</sup>, was derepressed in *SMRT*<sup>-/-</sup> neural stem cells (Fig. 4a) and efficiently induced by RA in neural stem cells (Fig. 4b) as well as in HeLa and U2OS cells (Supplementary Fig. 5e). ChIP experiments confirmed recruitment of SMRT and RA receptor to the canonical RARE region of the *JMJD3* promoter region (Fig. 4c, d), suggesting that *JMJD3* is a direct target of SMRT-dependent RA-receptor-mediated repression. Overexpression of *JMJD3* in neural stem cells resulted in induction of neuronal genes of various neuronal subtypes as assessed by quantitative (q)PCR (Fig. 4e). Overexpression of a mutant form of *JMJD3*, which lacks the enzymatic jumonji C domain, failed to activate neuronal genes in the neural stem cells (Fig. 4e), indicating that the catalytic functions of *JMJD3* permit the neurogenic program to unfold.

Many jumonji-C-domain containing proteins catalyse removal of methylated histone tails but nevertheless show distinct substrate specificity among various protein families<sup>20,21</sup>. *In vitro* histone demethylase assays revealed selective removal of H3K27me3 (Fig. 4f) and expression of full-length *JMJD3*, but not of jumonji-C-domain deleted protein, resulted in a dramatic decrease of global trimethyl histone H3K27 level (Fig. 4g), indicating that *JMJD3* demethylates H3K27me3. Other histone modifications evaluated, including H3K27me2, were not diminished by overexpression of *JMJD3*. Our data demonstrate that *JMJD3* is a functional demethylase that specifically targets histone H3K27me3 and this function has been independently reported for both *JMJD3* and the highly

homologous UTX protein (while this manuscript was under revision<sup>23–26</sup>).

Trimethyl K27 modification is important in transmitting epigenetic information during development, imprinting and X-chromosome inactivation and has limited reversibility<sup>27</sup>. A global 'passive' decrease of K27me3 has been observed in embryonic stem cell differentiation, accompanied by concurrent loss of the Polycomb group protein Ezh2 (ref. 28), the only known histone methyltransferase that targets K27. We observed, however, that Ezh2 levels remain constant after RA-stimulated differentiation of neural stem cells (Fig. 4b), prompting the suggestion that epigenetic regulation via histone H3K27me3 modification by the histone demethylase *JMJD3* is a contributory early event in neural differentiation. Indeed, RA-induced neuronal differentiation caused a decrease in histone H3K27me3 and recruitment of *JMJD3* to the *Dlx5* promoter (Fig. 4h, i). These results are consistent with several recent genome-wide location analyses that



**Figure 4 | Derepression of neurogenic genes promotes neuronal differentiation program.** **a**, mRNA levels of *JMJD3* in neural stem cells ( $\pm$ s.d.,  $^{**}P < 0.005$ ). **b**, mRNA levels of *EZH2* and *JMJD3* in the absence or presence of RA (10<sup>-6</sup> M) ( $\pm$ s.d.,  $^{**}P < 0.005$ ). **c**, ChIP of anti-SMRT on the *JMJD3* promoter ( $\pm$ s.d.). **d**, ChIP of  $\alpha$ -RA receptor on the *Jmjd3* promoter in E14.5 rat embryonic telencephalon ( $\pm$ s.d.,  $^{**}P < 0.005$ ). **e**, mRNA expression levels of neurogenic genes in neural stem cells after transfection with the indicated plasmids ( $\pm$ s.d.,  $^{*}P < 0.05$ ;  $^{**}P < 0.005$ ). **f**, *In vitro* demethylation assay using anti-Flag immunoprecipitated *JMJD3* incubated with calf thymus histones and immunoblotted using modification-specific antibodies. **g**, Whole-cell extracts from HEK cells transfected with *JMJD3* or *JMJD3mut* immunoblotted modification-specific antibodies. TBP, TATA-binding protein. **h**, ChIP analysis of H3K27me3 levels on the *Dlx5* promoter ( $\pm$ s.e.m.,  $^{*}P < 0.05$ ;  $^{**}P < 0.005$ ). IgG, immunoglobulin  $\gamma$ . **i**, ChIP of *JMJD3* on the *Dlx5* promoter in Flag-*JMJD3* transfected HEK cells using anti-Flag IgG ( $\pm$ s.d.,  $^{**}P < 0.005$ ).

have suggested histone H3K27me3 to be a dynamic marker reflecting developmental potential<sup>29–32</sup>.

Our results suggest that expression of JMJD3, a novel histone H3K27 demethylase, is regulated in neural stem cell differentiation in response to RA by SMRT-dependent, RA-receptor-mediated programs, and serves as a mechanistic component of this neuronal fate program. Together, our studies have uncovered a specific role of SMRT in maintaining the neural stem cell state, defending against an ability of unliganded RA receptor to initiate a differentiation program along a neuronal pathway based, at least in part, on the RA-dependent regulation of a component of the histone methylation/demethylation machinery.

## METHODS SUMMARY

**Cell culture and transfection.** Neural stem cells were isolated by dissecting E13 mouse cortices or E14.5 rat cortices, and cultured in the presence of 10 ng ml<sup>-1</sup> FGF2 as previously described<sup>7</sup>. NTERA2, a human embryonal carcinoma stem cell line and U2OS, a human osteosarcoma epithelial cell line, were maintained in  $\alpha$ -MEM medium supplemented with 10% FBS. Mouse neural stem cells were transfected using Lipofectamine 2000 (Invitrogen) and 72 h later were fixed for immunostaining. Transfection of rat neural stem cells was carried out with Amaxa Nucleofactor system (following the manufacturer's instructions) using 10<sup>6</sup> neural stem cells and 10  $\mu$ g various plasmid DNA for each nucleofection. After nucleofection using the A-31 program, neural stem cells were maintained in the presence of FGF2 (10 ng ml<sup>-1</sup>) for 48 h before harvesting RNA for quantitative RT-PCR analysis.

**In vitro demethylase assay.** Demethylase assay was performed as described<sup>33</sup>. Briefly, Flag-Jmjd3 was immunoprecipitated and partially purified from HEK293 cells and was incubated with calf thymus histones (Sigma) at 37 °C for 2–3 h in histone demethylation buffer: 50 mM HEPES-KOH (pH 7.5), 70  $\mu$ M Fe(NH<sub>4</sub>)<sub>2</sub>(SO<sub>4</sub>)<sub>2</sub>, 1 mM  $\alpha$ -ketoglutarate, and 2 mM ascorbate. Histone methylation levels were analysed by western blot with specific antibodies.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 30 July; accepted 20 September 2007.

Published online 10 October 2007.

- Guillemot, F. Cellular and molecular control of neurogenesis in the mammalian telencephalon. *Curr. Opin. Cell Biol.* **17**, 639–647 (2005).
- Hsieh, J. & Gage, F. H. Epigenetic control of neural stem cell fate. *Curr. Opin. Genet. Dev.* **14**, 461–469 (2004).
- Ross, S. E., Greenberg, M. E. & Stiles, C. D. Basic helix-loop-helix factors in cortical development. *Neuron* **39**, 13–25 (2003).
- Chen, J. D. & Evans, R. M. A transcriptional co-repressor that interacts with nuclear hormone receptors. *Nature* **377**, 454–457 (1995).
- Horlein, A. J. *et al.* Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor. *Nature* **377**, 397–404 (1995).
- Nishihara, E., O'Malley, B. W. & Xu, J. Nuclear receptor coregulators are new players in nervous system development and function. *Mol. Neurobiol.* **30**, 307–325 (2004).
- Hermanson, O., Jepsen, K. & Rosenfeld, M. G. N-CoR controls differentiation of neural stem cells into astrocytes. *Nature* **419**, 934–939 (2002).
- Jepsen, K. *et al.* Combinatorial roles of the nuclear receptor corepressor in transcription and development. *Cell* **102**, 753–763 (2000).
- Campbell, K. Dorsal-ventral patterning in the mammalian telencephalon. *Curr. Opin. Neurobiol.* **13**, 50–56 (2003).
- Akiyama, H. *et al.* The role of transcriptional corepressor Nif31 in early stage of neural differentiation via cooperation with Trip15/CSN2. *J. Biol. Chem.* **278**, 10752–10762 (2003).
- Anderson, S. A. *et al.* Mutations of the homeobox genes *Dlx-1* and *Dlx-2* disrupt the striatal subventricular zone and differentiation of late born striatal neurons. *Neuron* **19**, 27–37 (1997).
- Ribes, V., Wang, Z., Dolle, P. & Niederreither, K. Retinaldehyde dehydrogenase 2 (RALDH2)-mediated retinoic acid synthesis regulates early mouse embryonic forebrain development by controlling FGF and sonic hedgehog signaling. *Development* **133**, 351–361 (2006).
- Halilagic, A. *et al.* Retinoids control anterior and dorsal properties in the developing forebrain. *Dev. Biol.* **303**, 362–375 (2007).
- Park, J. I., Tsai, S. Y. & Tsai, M. J. Molecular mechanism of chicken ovalbumin upstream promoter-transcription factor (COUP-TF) actions. *Keio J. Med.* **52**, 174–181 (2003).
- Cohen, R. N., Putney, A., Wondisford, F. E. & Hollenberg, A. N. The nuclear corepressors recognize distinct nuclear receptor complexes. *Molec. Endocrinol.* **14**, 900–914 (2000).
- Kao, H. Y. *et al.* A histone deacetylase corepressor complex regulates the Notch signal transduction pathway. *Genes Dev.* **12**, 2269–2277 (1998).
- Stancheva, I., Collins, A. L., Van den Veyver, I. B., Zoghbi, H. & Meehan, R. R. A mutant form of MeCP2 protein associated with human Rett syndrome cannot be displaced from methylated DNA by Notch in *Xenopus* embryos. *Mol. Cell* **12**, 425–435 (2003).
- Tsuda, L., Nagaraj, R., Zipursky, S. L. & Banerjee, U. An EGFR/Ebi/Sno pathway promotes Delta expression by inactivating Su(H)/SMRTER repression during inductive Notch signaling. *Cell* **110**, 625–637 (2002).
- Rossant, J., Zirngibl, R., Cado, D., Shago, M. & Giguere, V. Expression of a retinoic acid response element-*hsp1acZ* transgene defines specific domains of transcriptional activity during mouse embryogenesis. *Genes Dev.* **5**, 1333–1344 (1991).
- Klose, R. J., Kallin, E. M. & Zhang, Y. JmJC-domain-containing proteins and histone demethylation. *Nature Rev. Genet.* **7**, 715–727 (2006).
- Shi, Y. & Whetstone, J. R. Dynamic regulation of histone lysine methylation by demethylases. *Mol. Cell* **25**, 1–14 (2007).
- Treweek, S. C., McLaughlin, P. J. & Allshire, R. C. Methylation: lost in hydroxylation? *EMBO Rep.* **6**, 315–320 (2005).
- Agger, K. *et al.* UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature advance online publication*, doi:10.1038/nature06145 (22 August 2007).
- De Santa, F. *et al.* The histone H3 lysine-27 demethylase Jmjd3 links inflammation to inhibition of polycomb-mediated gene silencing. *Cell* **130**, 1083–1094 (2007).
- Lee, M. G. *et al.* Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. *Science Express* doi:10.1126/science.1149042 (30 August 2007).
- Lan, F. *et al.* A histone H3 lysine 27 demethylase regulates animal posterior development. *Nature advance online publication*, doi:10.1038/nature06192 (12 September 2007).
- Trojer, P. & Reinberg, D. Histone lysine demethylases and their impact on epigenetics. *Cell* **125**, 213–217 (2006).
- Lee, E. R., Murdoch, F. E. & Fritsch, M. K. High histone acetylation and decreased polycomb repressive complex 2 member levels regulate gene specific transcriptional changes during early embryonic stem cell differentiation induced by retinoic acid. *Stem Cells advance online publication*, doi:10.1634/stemcells.2007-0203 (24 May 2007).
- Bracken, A. P., Dietrich, N., Pasini, D., Hansen, K. H. & Helin, K. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev.* **20**, 1123–1136 (2006).
- Chang, S. & Aune, T. M. Dynamic changes in histone-methylation 'marks' across the locus encoding interferon-gamma during the differentiation of T helper type 2 cells. *Nature Immunol.* **8**, 723–731 (2007).
- Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007); published online 1 July 2007.
- Roh, T. Y., Cuddapah, S. & Zhao, K. Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev.* **19**, 542–552 (2005).
- Tsukada, Y. *et al.* Histone demethylation by a family of JmJC domain-containing proteins. *Nature* **439**, 811–816 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank H. Taylor for animal care; C. Nelson for various cell culture assistance; J. Hightower for artwork; M. Fisher for assistance with the manuscript; J. Rossant for providing RARE-LacZ mice; F. Liu for ES cell injections; and the UCSD BIOGEM laboratory (G. Hardiman, director) for array processing and bioinformatics analysis. BIOGEM was supported by an NIH/NIDDK award. M.G.R. is an investigator with the Howard Hughes Medical Institute. H.-J.K. is supported by a post-doctoral fellowship from the Susan G. Komen Foundation. K.J. is supported by a Scientist Development Grant from the American Heart Institute. T.Z. is supported by a post-doctoral fellowship from PCRP of CDMRP. O.H. is supported by grants from VR, CF, BCF and SSF. This work was funded by grants from the NIH to M.G.R.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to K.J. (kjepsen@ucsd.edu).



## METHODS

**Targeting construct, gene targeting, and generation of knockout mice.** Genomic DNA clones encoding the mouse *SMRT* gene were isolated from a Lambda FIXII library made from the 129Sv mouse strain (Stratagene), using a mouse *SMRT* cDNA fragment of residues 3,090–4,035 as a probe. The targeting construct was designed to delete four exons containing amino acids 1,038 to 1,198. The targeting vector contained the splice acceptor from the third exon of the rabbit  $\beta$ -globin gene to help prevent random splicing. Genotypes were determined by PCR analysis using the following primer pairs: *SMRT* wild-type 5': GATGTCAGTCCAGCTTCGTG; *SMRT* wild-type 3': CCTTG-GTGATGCTT-CCACTG; *SMRT* recombined 5': CTCACCTCGGAAGGAC-ATATG; *SMRT* recombined 3': CTAGTGAGACGTGCTACTTC. 35 cycles of PCR with an annealing temperature of 60 °C was performed. The 5' probe used for both Southern and northern blots corresponds to *SMRT* base pairs 2,760–3,209, while the 3' probe used for the Southern blot corresponds to base pairs 3,114–3,261. Transgenic mice expressing full-length *SMRT* under the  $\alpha$ -MHC promoter were generated and bred onto a *SMRT*-null genetic background.

**Histology.** Frozen formalin-fixed tissue sections were prepared for *in situ* hybridization and H&E staining as previously described<sup>34</sup>. Probes for *in situ* analysis were generated by PCR amplifications of 3' UTR exons. Immunostaining was performed following standard protocols on cells fixed in 4% paraformaldehyde/PBS, using AlexaFluor-conjugated secondary antibodies (Molecular Probes).

**Reagents and antibodies.** The Flag-JMJD3 construct was cloned by inserting full-length mouse *Jmjd3* cDNA in frame into p3x FLAG CMV-10 (Sigma) vector within *HindIII* and *BamHI* sites. The Flag-JMJD3mut was generated by removing the carboxy-terminus 410 amino acids, which include the jumonji C domain. The following antibodies were obtained from Upstate Biotechnology: anti-H3K27me2 (07-031), anti-H3K27me3 (07-449), anti-H3K20me3 (07-463), anti-H3K4me3 (07-473), anti-H3K9me3 (07-442), anti-H3K4me2 (07-030), anti-H3K36me3 (07-549), anti-H3K4me2 (07-030), anti-pan-histone (06-755). Additionally, we used anti-H3K79me3 (Abcam, ab2621), mouse anti-Flag M2 (Sigma) and anti-TBP (Santa Cruz, sc-33736), rabbit anti-NCoR<sup>34</sup>, rabbit anti-*SMRT* (ABR, PAI-842), mouse anti-TuJ1 (Babco), rabbit anti-GFAP (Chemicon), nestin (DSHB), and anti- $\beta$ -Gal (Cappel), Ki67 (Novus) and cleaved caspase-3 (Cell Signalling Technologies) antibodies.

**Immunoprecipitation and immunoblotting.** For immunoprecipitation and immunoblotting, cells were rinsed in PBS, harvested and lysed in NETN buffer in the presence of protease inhibitors (Roche) and 1 mM phenylmethylsulphonylfluoride (PMSF). After sonicating briefly and clearing by centrifugation, the concentration of the extracts was determined using a Bradford protein assay kit (Sigma) and 20  $\mu$ g extracts were boiled in SDS sample buffer and loaded directly onto a 4–12% bis-tris NuPAGE gel (Invitrogen), transferred onto PVDF membranes (Millipore) and western blotting was performed following standard protocols. For immunoprecipitation, extracts were incubated with anti-Flag antibody (Sigma) overnight at 4 °C, followed by incubation with protein A/G agarose beads (Santa Cruz Biotech), washed extensively, and proceeded to *in vitro* enzymatic activity assay.

**Chromatin immunoprecipitation.** Chromatin immunoprecipitation was performed as described<sup>35,36</sup>. Essentially, HEK and NTera2 cells were grown in DMEM or  $\alpha$ -MEM respectively with 10% charcoal-stripped medium for 48–72 h and then treated with 1  $\mu$ M retinoic acid for different times. Embryonic telecephalons were dissected and dispersed mechanically with cell strainer (BD Bioscience). Cells were crosslinked with 1% formaldehyde and the extracts were sonicated until the DNA fragments were 300–800 base pairs in average size. Crosslinked material was immunoprecipitated with 1–5  $\mu$ g of the specific IgGs overnight at 4 °C and followed by incubation with protein A-sepharose beads (Sigma) for an additional 2 h. After extensive washes, the isolated complexes were eluted and the crosslinking was reversed by heating at 65 °C overnight. Immunoprecipitated DNA was recovered by purification on QIAquick spin columns (Qiagen) and analysed by 25–30 cycles of PCR. The primers we used are: hDlx5-F, TCCTTCTTGCTCCATCCTA; hDlx5-R, GGGGACAATGTG-TTCCAATC; hJD3-f, CCAGGAGAGAGAAGGCACAG; hJD3-R, AGAAACCC-AGGCGTCCTTAT; rDCX-F, GCTTGATGGAGAGAGGGTTTGA; rDCX-R,

CCAAGGTTAGCATCTCCAGCAT; rDLL1-F, GCTTTGCGATAGTGGAAGAG; rDLL1-R, GAAGAGAGCCCAGATGTTCA; rDLX5-F, TAGGCAATGACAAA-GTATTCAGGA; rDLX5-R, TCTCAAGACTCCACAGAGATGTTTC; rgad1-F, CGAACGAGGTAGTTTGGAGA; rgad1-R, CCCGTATCTTGGAGGAGATT; rGFAP-f, GTTCCCAGGAAACCCCTTGACTC; rGFAP-R, GTGATACGTCTC-CGCTCCATCT; rHES5-F, GTTCCAGGTGGAGGGACACTCT; rHES5-R, GCATTTTGTGTGTGGGAGGTA; rJD3-F, ATGACCGGATTCCTAGGAT; rJD3-R, ACGTGCCTGGTGACATCAT; rNKX6.2-F, TCCCTGCCTTCTTT-CTCTCTGA; rNKX6.2-R, GCCGGCTCGGGAAGTTTG; rOLIG2-F, CCTCCA-CCCAGCTATAAAAACCAG; and rOLIG2-R, CCTAAGTCCTAGCCGAGTT-GGAAA.

**RNA isolation and RT-PCR analysis.** Cells were lysed directly on culture dishes in the presence of a highly denaturing buffer (RNeasy kit, Qiagen). The lysate, collected with a cell scraper, was pipetted directly onto a QIAshredder spin column (Qiagen) to shear the genomic DNA. Ethanol was added to adjust the binding conditions, and the sample was applied to an RNeasy spin column for binding of total RNA, following the manufacturer's protocol (Qiagen). First-strand cDNA synthesis from total RNA was carried out with mixture of oligodT and random primer using SuperscriptIII (Qiagen). The resulting cDNA was analysed by quantitative PCR (Stratagene MX3000p) using the following oligos: mJmjd3-F, GGACTCCGTCTGCTATCTGC; mJmjd3-R, ATGACCGGATTC-CCTAGGAT; mNIF3L1-F, GGAGAACATTAGAGTCCCAAGTCA; mNIF3L1-R, TCAAAGTGAACACCCAGCATTTTC; rUTL2-F, CATGGAGCAAGCTGA-CACAG; rUTL2-R, GAGGTGATGGTGAGGACGTT; rDcx-F, GCAAGCTC-AAATCTTCAGTTACAG; rDcx-R, TCGTTCGTCAAAATGTCCAA; rDLX5-F, CTGGCCCGCTTTACAGAGAAG; rDLX5-R, CTGGTGACTGTGGCGAGTTA; rEmx1-F, ACTTCTACCCCTGGGTGCTT; rEmx1-R, GCACCCACCACGTAG-TGATT; rEMX2-F, ACCTTCTACCCCTGGCTCAT; rEMX2-R, TAAAGCT-GGGACGGAGAGA; rEZH2-F, GCCATTTCTCAATGTTTCCAG; rEZH2-R, TCGTCGACAAAAGAGCGTATGA; rFOX2-F, CTTTCTAAGGAACGCGAA-CG; rFOX2-R, ACTGGCTGGGGTGATTACTG; rFz2-F, CAAGGTGTTCA-ATGCTCACTATAAC; rFz2-R, GCCGCACTGGTTACACTTATG; rGad1-F, GTGAGTGCCTTCAGGGAGAG; rGad1-R, CGTCTTGCGGACATAGTTGA; rGad2-F, CTGTGTACGGGGCTTTTGAT; rGad2-R, TGCATCAGTCCCTCC-TCTCT; rGUSB-F, GCCAATGAGCCTGTCTCTTC; rGUSB-R, CTCCAGAT-GCCCGTAGTCAT; rJmjd3-F, AATGATCAAGTTCTGCCTCCTG; rJmjd3-R, TGTTGAACACCTCCACATCG; rNEFL-F, GGACAAGCAGAATGCAGACA; rNEFL-R, CTACCCACGCTGGTGAAACT; rNif3L1-F, CTTTCTGCGCAGG-TGTGAT; rNif3L1-R, CAAGCGTCCCATTCAGTAT; rNKX2.2-F, GCCAT-CTCCCGCTTCATG; rNKX2.2-R, CGCGGACAGGTAATTCTGCT; rOtx1-F, GTTCGAAAAGACTCGTACC; rOtx1-R, CCGGAGACGACTTCTTCTTG; rPvalb-F, CCGGATGATGTGAAGAAGGT; and rPvalb-R, TTCTCAACCCC-AATCTTGC.

**RNA profiling.** RNA was prepared from individually cultured cells of E12.5 cortex from wild-type and *SMRT*<sup>-/-</sup> embryos 10 h after plating or RA treatment. RNA quality was assessed using the Agilent Bioanalyser 6000 Pico LabChip. 100 ng of total RNA was labelled with Cy-3 or Cy-5 using the Agilent Low RNA Input Fluorescent Linear Amplification Kit. A dye-swap design was employed. Labelled cDNA was hybridized to the Agilent 44K Whole Mouse Genome Array. Data was collected using the Agilent Microarray Scanner and Feature Extraction Software, using a Lowess option with spatial detrend to extract genes of interest with more confidence than through the use of fold-change only. Experiments were performed in triplicate, with litter-matched wild-type and mutant samples.

34. Jepsen, K. *et al.* Combinatorial roles of the nuclear receptor corepressor in transcription and development. *Cell* **102**, 753–763 (2000).
35. Perissi, V., Aggarwal, A., Glass, C. K., Rose, D. W. & Rosenfeld, M. G. A corepressor/coactivator exchange complex required for transcriptional activation by nuclear receptors and other regulated transcription factors. *Cell* **116**, 511–526 (2004).
36. Shang, Y., Hu, X., DiRenzo, J., Lazar, M. A. & Brown, M. Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell* **103**, 843–852 (2000).

## LETTERS

# Neural substrates of awakening probed with optogenetic control of hypocretin neurons

Antoine R. Adamantidis<sup>1\*</sup>, Feng Zhang<sup>2\*</sup>, Alexander M. Aravanis<sup>2</sup>, Karl Deisseroth<sup>1,2</sup> & Luis de Lecea<sup>1</sup>

The neural underpinnings of sleep involve interactions between sleep-promoting areas such as the anterior hypothalamus, and arousal systems located in the posterior hypothalamus, the basal forebrain and the brainstem<sup>1,2</sup>. Hypocretin<sup>3</sup> (Hcrt, also known as orexin<sup>4</sup>)-producing neurons in the lateral hypothalamus<sup>5</sup> are important for arousal stability<sup>2</sup>, and loss of Hcrt function has been linked to narcolepsy<sup>6–9</sup>. However, it is unknown whether electrical activity arising from Hcrt neurons is sufficient to drive awakening from sleep states or is simply correlated with it. Here we directly probed the impact of Hcrt neuron activity on sleep state transitions with *in vivo* neural photostimulation<sup>10–18</sup>, genetically targeting channelrhodopsin-2 to Hcrt cells and using an optical fibre to deliver light deep in the brain, directly into the lateral hypothalamus, of freely moving mice. We found that direct, selective, optogenetic photostimulation of Hcrt neurons increased the probability of transition to wakefulness from either slow wave sleep or rapid eye movement sleep. Notably, photostimulation using 5–30 Hz light pulse trains reduced latency to wakefulness, whereas 1 Hz trains did not. This study establishes a causal relationship between frequency-dependent activity of a genetically defined neural cell type and a specific mammalian behaviour central to clinical conditions and neurobehavioural physiology.

To target genetically expression of channelrhodopsin-2 (ChR2) to Hcrt neurons, we used a lentivirus carrying the 3.1-kilobase (kb) mouse prepro-hypocretin (*Hcrt*, which encodes hypocretin 1 and hypocretin 2) gene promoter<sup>19,20</sup> (Fig. 1a). The specificity of lentivirus-mediated expression was tested by stereotactic delivery of concentrated *Hcrt::ChR2-mCherry* lentiviruses ( $>10^9$  plaque-forming units (p.f.u.) ml<sup>-1</sup>) into the lateral hypothalamus of *Hcrt::EGFP* transgenic mice<sup>21</sup>. Expression persisted for at least 2 months (Fig. 1b). Out of 748 total enhanced green fluorescent protein (EGFP)-expressing Hcrt cells in the targeted lateral hypothalamus of transduced animals, 658 cells were positive for ChR2-mCherry ( $87.9 \pm 0.9\%$ , Fig. 1c) and  $\sim 97\%$  of mCherry-expressing cells also clearly co-expressed EGFP ( $96.9 \pm 1.3\%$ ,  $n = 3$  animals; Fig. 1c).

To test whether ChR2-expressing Hcrt neurons could precisely respond to light pulses, we used the whole-cell voltage-clamp technique to measure photocurrents in acute lateral hypothalamus brain slices from *Hcrt::EGFP* transgenic mice<sup>21</sup> injected with *Hcrt::ChR2-mCherry* lentiviruses. Without the addition of any exogenous all-*trans*-retinal<sup>12</sup>, all of the cells co-expressing ChR2-mCherry (red) and EGFP (green) in the hypothalamus exhibited photocurrents with a peak amplitude of  $182 \pm 86$  pA (mean  $\pm$  s.d.,  $n = 7$  cells; example photocurrent shown in Fig. 1d). In contrast, the occasional EGFP-negative neurons that exhibited dim red fluorescence after virus injection did not respond to blue light under whole-cell patch clamp ( $n = 5$  out of 5 cells from 3 animals). In the cells co-expressing EGFP and

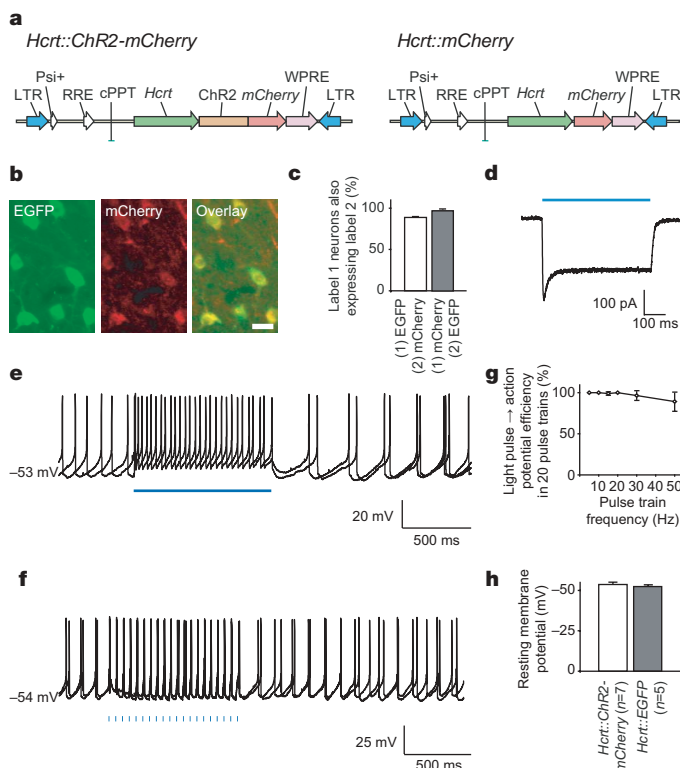
mCherry, continuous light illumination for 1 s evoked robust action potential trains under current-clamp conditions (Fig. 1e) with firing frequencies ranging from 8 to 25 Hz. Moreover, as in hippocampal neurons<sup>12</sup>, 15-ms blue-light pulses were able to drive reliable action potential trains from 1 to 50 Hz under whole-cell current-clamp conditions (Fig. 1f), although reliability decreased slightly with increasing frequency (Fig. 1g). At 20 Hz, 100% of light pulses gave rise to action potentials, whereas at 50 Hz,  $84 \pm 14\%$  of light pulses gave rise to action potentials ( $n = 6$  cells; Fig. 1g). We found that basal properties including input resistance ( $366 \pm 43$  M $\Omega$ ,  $n = 6$ )<sup>21</sup> and resting membrane potential were unaltered by ChR2-mCherry expression ( $P > 0.05$ ; Fig. 1h). Together, these data show that lentivirus-mediated expression of ChR2-mCherry in Hcrt neurons is well tolerated, specific and sufficient to drive precise and reliable action potential firing across a range of frequencies.

We next studied the effect of photostimulation of *Hcrt::ChR2-mCherry* neurons on the sleep–wake cycle of freely moving mice. *Hcrt::ChR2-mCherry* or *Hcrt::mCherry* (control) lentiviruses were stereotactically delivered to the lateral hypothalamus; to determine the behavioural consequences of Hcrt neuron activation, we quantitatively analysed sleep recordings of mice chronically implanted with electroencephalographic (EEG) and electromyographic (EMG) electrodes (Fig. 2a). Under baseline conditions, the spontaneous sleep–wake cycle of *Hcrt::ChR2-mCherry* transduced animals was not different from *Hcrt::mCherry* control animals ( $P > 0.05$ ,  $n = 4$  animals; Supplementary Fig. 1). We coupled a 200- $\mu$ m optical fibre to a 20-mW 488-nm diode laser to deliver light into the lateral hypothalamus (Fig. 2a), and fed the optical fibre through the same cannula guide used to deliver the lentiviruses. We measured light scattering in mouse lateral hypothalamus tissue slices (Fig. 2b, c), and estimated that a volume of 0.5 mm<sup>3</sup> received at least 1 mW mm<sup>-2</sup> (required for ChR2 activation<sup>10</sup>), consistent with the estimated volume of the Hcrt neuronal field ( $\sim 0.56$  mm<sup>3</sup>) (Fig. 2c). We found that *in vivo* photostimulation (pulse trains (10 s, 20 Hz) delivered once per minute over 10 min) significantly increased the percentage of Hcrt neurons expressing the c-Fos protein (a marker of neuronal activity) in *Hcrt::ChR2-mCherry* animals compared to the *Hcrt::mCherry* control animals ( $65.26 \pm 4.29\%$  versus  $25.72 \pm 5.38\%$ , respectively,  $P < 0.01$ ; see Supplementary Fig. 3). This result suggests that *in vivo* photostimulation efficiently activates the Hcrt neuron population.

To test systematically for a causal role of Hcrt neurons in awakening, we next quantitatively measured latency of sleep-to-wake transitions from either slow wave sleep (SWS) or rapid eye movement (REM) sleep in animals transduced with either *Hcrt::ChR2-mCherry* or control lentiviruses (Fig. 3a). On the basis of previous electrical recording of Hcrt neuron activity *in vivo*<sup>22,23</sup>, we stimulated each animal with 10 s of light illumination, either in the form of pulsed light flashes (15-ms pulse width) ranging from 1 to 30 Hz,

<sup>1</sup>Department of Psychiatry and Behavioral Sciences, Stanford University, 701B Welch Road, Palo Alto, California 94304, USA. <sup>2</sup>Department of Bioengineering, Stanford University, James H. Clark Center W083, Stanford, California 94305, USA.

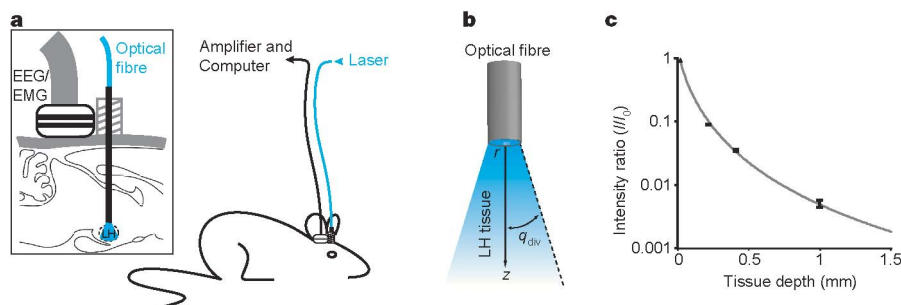
\*These authors contributed equally to this work.



**Figure 1 | Genetically targeted cell-type-specific optical control of Hcrt neurons using ChR2.** **a**, Schematic diagram showing lentiviral vectors carrying the genes for ChR2-mCherry or mCherry driven by the 3.1-kb mouse *Hcrt* promoter<sup>19,20</sup>. The lentiviral backbone is derived from pLenti-CaMKII $\alpha$ -ChR2-mCherry<sup>25</sup>. LTR, long terminal repeats; RRE, Rev responsive element; WPRE, woodchuck post-transcriptional regulatory element. **b**, Images of Hcrt neurons (green, *Hcrt::EGFP*) co-expressing ChR2-mCherry protein (red) in the adult mouse lateral hypothalamus (scale bar, 20  $\mu$ m). Hcrt and mCherry expression was detected by double-fluorescent immunohistochemistry (see Methods). **c**, Lentiviral *Hcrt::ChR2-mCherry* expression is highly specific to the *Hcrt::EGFP* neurons. **d**, *Hcrt::ChR2-mCherry* neurons in the lateral hypothalamus can be electrically controlled with light. Voltage-clamp recording of a neuron expressing ChR2-mCherry in acute lateral hypothalamus slice shows inward photocurrent evoked by illumination with blue light. **e**, Neurons expressing *Hcrt::ChR2-mCherry* in acute lateral hypothalamus brain slices under current-clamp conditions fire action potentials on illumination with 1 s of continuous blue light. Two sweeps are superimposed. **f**, Blue-light pulse trains (15 ms per pulse, 20 Hz) evoked reliable firing of action potential trains. Two consecutive sweeps are superimposed, showing temporal precision of evoked action potential trains even in the presence of basal spontaneous activity (spontaneous activity in Hcrt neurons *in vitro* is expected, as previously reported<sup>29</sup>). Fifteen-millisecond light pulses are indicated by blue bars. **g**, Light-evoked spike trains are reliable over a range of frequencies. The percentage of action potentials evoked by 20 light pulses at the indicated frequency (15–50 Hz) is shown ( $n = 6$ ). **h**, Comparison between the resting membrane potential of ChR2-mCherry/EGFP double-positive ( $n = 7$ ) and EGFP-only ( $n = 5$ ) neurons; ChR2-mCherry expression does not significantly alter basal electrical properties of the Hcrt neurons. Error bars, s.e.m.

or 10 s of continuous illumination. To quantify the effect of each stimulation frequency, we measured the latency between the end of the photostimulation and the next transition to wakefulness (see Methods for description of the sleep–wake cycle criteria used; Fig. 3b, c). Light pulses at 1 Hz did not result in a significant change in latency of SWS or REM sleep to wakefulness in *Hcrt::ChR2-mCherry* animals compared to their controls or their unstimulated baseline recording ( $P > 0.05$ ; Fig. 3b, c). However, photostimulation of Hcrt neurons at frequencies between 5 and 30 Hz during SWS sleep markedly reduced the latency to wakefulness of *Hcrt::ChR2-mCherry* animals compared to control animals ( $P < 0.0001$  for all comparisons with control; Fig. 3b and Supplementary Movie 1). Ten-second continuous light stimulation (ON), expected from our *in vitro* work to give rise to  $> 5$  Hz firing of the Hcrt neurons, elicited a similar effect ( $P < 0.0001$ ; Fig. 3b). We also systematically studied the

effect of different stimulation frequencies on the latency of REM sleep-to-wake transitions, finding that 5–20 Hz stimulation and constant illumination robustly decreased the latency to wakefulness in *Hcrt::ChR2-mCherry* animals compared to control animals ( $P < 0.05$ ; Fig. 3c and Supplementary Movie 2). Photostimulation effects were consistent not only across frequencies but also across animals, as pooled wake latency values from both SWS and REM sleep for each animal revealed significant differences between experimental and control animals (*Hcrt::mCherry* versus *Hcrt::ChR2-mCherry*: SWS sleep,  $57.27 \pm 2.67$  s and  $33.11 \pm 2.12$  s, respectively (Fig. 4c); REM sleep,  $57.90 \pm 3.25$  s and  $30.29 \pm 2.88$  s, respectively;  $P < 0.05$ , Student's *t*-test). The Hcrt-neuron-driven latencies to awakening were not fixed; instead, photostimulation shifted the distribution of sleep-to-wake latencies (Fig. 3d, e). Exemplar cumulative probability histograms are shown for 20 Hz and 1 Hz superimposed



**Figure 2 | Integrated *in vivo* optical and physiological system for control of the lateral hypothalamus in the setting of behavioural analysis.** **a**, Schematic of the behavioural set-up used for *in vivo* deep-brain photostimulation in mice. Magnification (inset) shows the EEG/EMG connector used for sleep recording and the cannula guide used for lateral hypothalamus light delivery through an optical fibre. **b**, Schematic of experimental set-up, showing relationship between the optical fibre, brain tissue and attenuating light.  $r$ , optical fibre radius;  $z$ , tissue depth from fibre end;  $q_{div}$ , the half-angle of divergence. LH, lateral hypothalamus. **c**, Normalized light intensity ( $mW mm^{-2}$ ) as a function of lateral hypothalamus tissue depth  $z$ . Values were experimentally determined as previously described<sup>14</sup> by measuring the

light intensity after transmission through a given tissue thickness and dividing by the intensity of the light emanating from the optical fibre tip. Tissue of different thickness was prepared in the form of acute brain slices from adult C57BL/6 mice. Error bars indicate one standard deviation from the mean. Sample size: 0.2 mm,  $n = 8$ ; 0.4 mm,  $n = 6$ ; 1 mm,  $n = 2$ . Fits were produced using the Kubelka–Munk model of light transmission through diffuse scattering media as previously described<sup>14</sup>. We estimate that by placing the tip of the optical fibre at the upper limit of the lateral hypothalamus, at least  $1 mW mm^{-2}$  of light, which is sufficient to activate ChR2<sup>10</sup>, reaches the entire Hcrt field ( $0.75 mm$  (medial-lateral)  $\times$   $0.75 mm$  (dorso-ventral)  $\times$   $1 mm$  (rostral-caudal)).

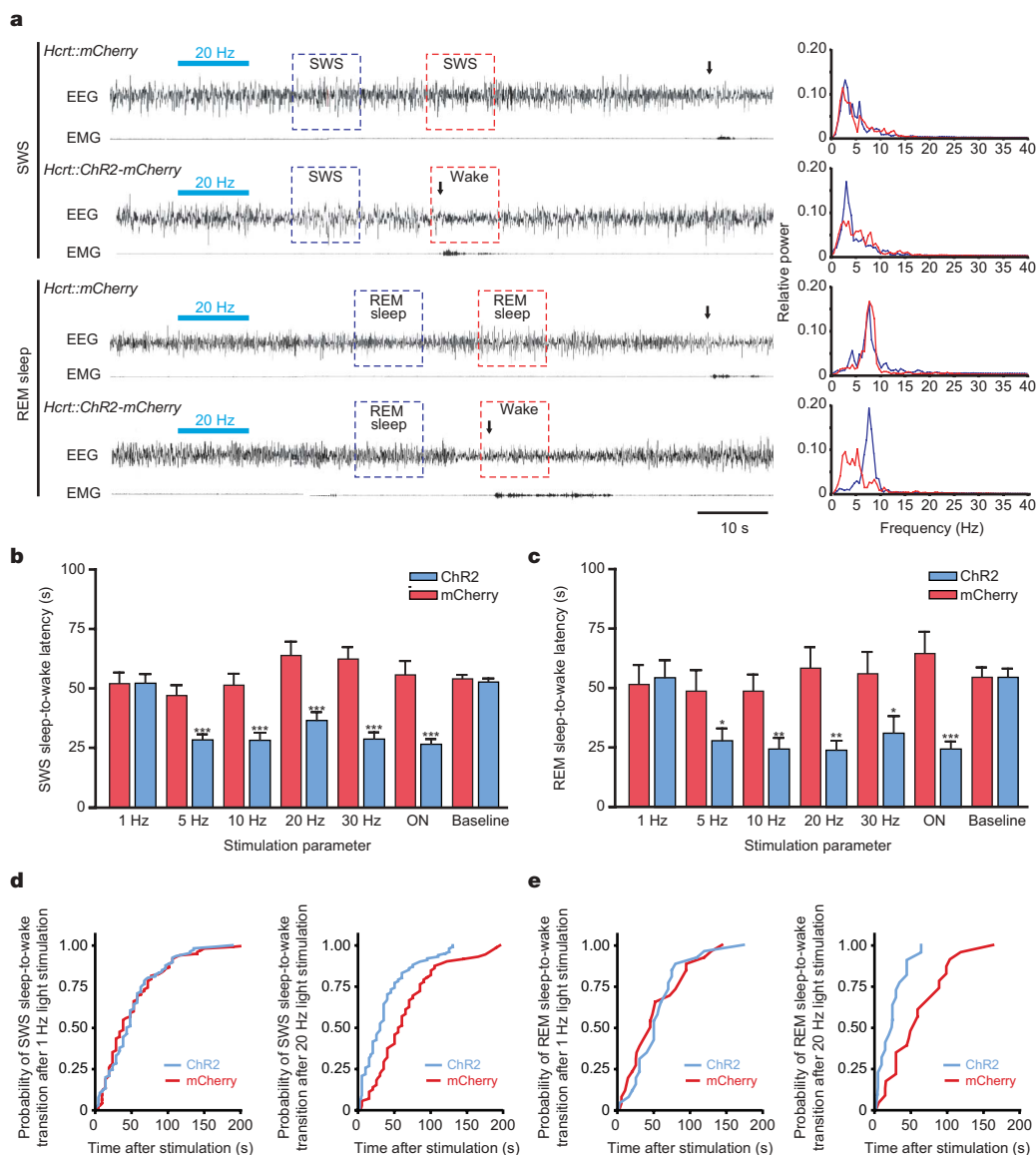


on respective control distributions, indicating that specific activation of Hcrt neurons increases the probability of awakening from either SWS (Fig. 3d) or REM sleep (Fig. 3e).

Effects of photostimulation seemed to be limited to the sleep-to-wake transition, as durations of the waking events after light delivery were not significantly different between *Hcrt::ChR2-mCherry* and *Hcrt::mCherry* control animals ( $P > 0.05$  for each frequency; data not shown). To test further whether Hcrt neuron stimulation specifically modulated arousal state transitions, we developed a semi-chronic photostimulation protocol (pulse trains (10 s, 20 Hz) delivered once per minute over 1 h), and quantified behavioural state transitions and durations during the entire period. We found that this stimulation paradigm increased the number of SWS sleep-to-wake

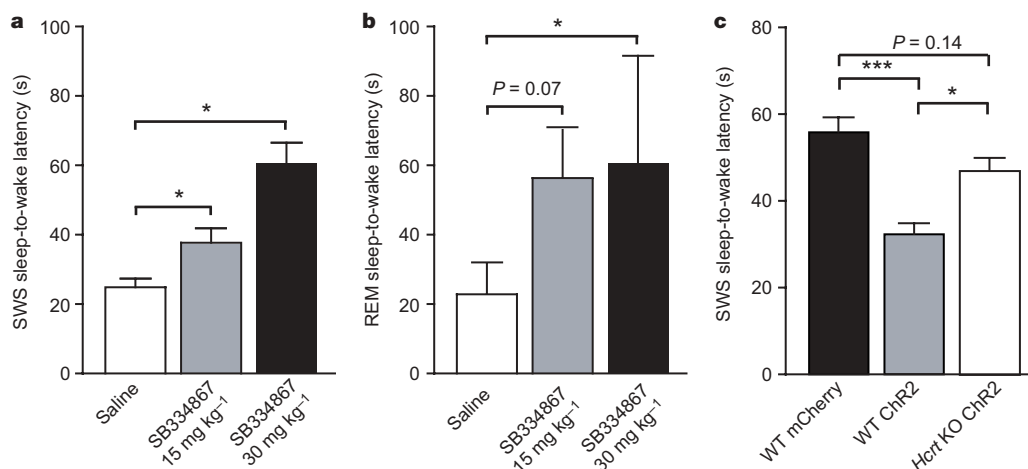
transitions in *Hcrt::ChR2-mCherry* animals compared to the control animals ( $P < 0.05$ , two-way analysis of variance (ANOVA) followed by Bonferroni post-hoc tests; Supplementary Table 1) but not total wake, SWS and REM amounts ( $P > 0.05$ , two-way ANOVA; Supplementary Table 2) or total EEG power characteristics ( $P > 0.05$ , two-way ANOVA; Supplementary Fig. 2a, b). The relative rarity of REM sleep events (due to the increase in SWS sleep-to-wake transitions) was probably responsible for the absence of change in REM sleep-to-wake transitions in *Hcrt::ChR2-mCherry* animals compared to the control animals in this protocol, as photostimulation was delivered at fixed intervals, independent of vigilance states.

Do these genetically targeted and photosensitized neurons influence neural circuitry via Hcrt release? We administered the Hcrt



**Figure 3 | In vivo photostimulation of Hcrt neurons drives sleep-to-wake transitions.** **a**, Representative EEG/EMG recordings showing awakenings after a single bout of photostimulation (15 ms, 20 Hz, 10 s) in *Hcrt::ChR2-mCherry* and *Hcrt::mCherry* (control) animals during SWS (upper traces) and REM sleep (lower traces). Light stimulations are represented by horizontal blue bars. Awakening events are indicated by vertical black arrows according to the described criteria (see Methods). Panels (right) show representative relative cortical EEG power spectra corresponding to the SWS and REM sleep-to-wake transitions highlighted with boxes on the EEG traces (left). **b**, **c**, Latencies of wake transitions during SWS (**b**) and REM sleep (**c**) of *Hcrt::ChR2-mCherry* transduced animals ( $n = 7$ ) and their controls ( $n = 6$ ) after a single photostimulation bout at different frequencies (15-ms light pulses, at 1–30 Hz,

during 10 s; ON, continuous light illumination of 10 s). Data analysis is based on an average of 15 and 5 stimulations per frequency and per mouse during SWS and REM sleep, respectively. Paired comparison between control conditions for SWS and REM sleep-to-wake transitions did not reveal any significant differences ( $P > 0.05$ , two-tailed Student's  $t$ -test). Latencies are represented as mean  $\pm$  s.e.m. Asterisk,  $P < 0.05$ ; double asterisk,  $P < 0.001$ ; triple asterisk,  $P < 0.0001$  using a two-tailed Student's  $t$ -test between mCherry control (red) and ChR2 animals (blue) for each frequency. **d**, Cumulative probability distribution of latencies from SWS to wakefulness after light stimulation (mCherry control, red curve; ChR2, blue curve). **e**, Cumulative probability distribution of latencies from REM sleep to wakefulness after light stimulation (mCherry control, red curve; ChR2, blue curve).



**Figure 4 | Behavioural transitions induced by photostimulation are mediated by Hcrt.** **a, b**, Effect of the Hcrt receptor 1 antagonist SB334867 (ref. 24) on latencies of light-induced wake events during SWS (**a**) and REM sleep (**b**) in *Hcrt::ChR2-mCherry* transduced animals ( $n = 3$ ) and their controls ( $n = 3$ ) after single bouts of 20 Hz photostimulation (15 ms, 10 s). Data analysis is based on an average of ten and three stimulations per frequency and per mouse for SWS and REM sleep, respectively. At the doses tested here, SB334867 had no effect on the latency of SWS and REM sleep-to-wake transitions after photostimulation in *Hcrt::mCherry* control animals. Latencies are represented as mean  $\pm$  s.e.m. Asterisk,  $P < 0.05$  using a paired Student's *t*-test between saline and drug conditions. **c**, Latencies of SWS sleep-to-wake transitions of *Hcrt* knockout (KO) animals compared with our

control data; three independent mice were transduced with *Hcrt::ChR2-mCherry* lentiviruses and received a single photostimulation bout spanning the relevant frequencies (1–30 Hz, 15-ms light pulses, during 10 s). *Hcrt* knockout data analysis is based on an average of 20 stimulations at different frequencies (1–30 Hz) per mouse during SWS. Latencies are represented as pooled mean  $\pm$  s.e.m. values. Triple asterisk,  $P < 0.0001$ , indicates the significant difference between control mCherry and ChR2 animals using a two-tailed Student's *t*-test. Asterisk,  $P < 0.05$ , indicates significant difference using a two-tailed Student's *t*-test between ChR2 and *Hcrt* knockout ChR2 animals. No significant differences in latencies to wakefulness were found between control mCherry and *Hcrt* knockout ChR2 animals.

receptor 1 antagonist SB334867 (ref. 24) 45 min before EEG/EMG recordings and *in vivo* photostimulation with 10-s trains of 20 Hz, 15-ms light pulses. As expected, 20 Hz photostimulation gave rise to brief-latency SWS and REM sleep-to-wake transitions in vehicle-treated animals (Fig. 4a, b). However, administration of a single dose of SB334867 (30 mg kg<sup>-1</sup>) blocked the photostimulation effect for both SWS and REM sleep-to-wake transitions (Fig. 4a, b;  $P < 0.05$ , paired *t*-test). A lower dose of the drug (15 mg kg<sup>-1</sup>) had a similar but reduced effect ( $P < 0.05$  for SWS sleep-to-wake transitions;  $P = 0.07$  for REM sleep-to-wake transitions, paired Student's *t*-test; Fig. 4a, b, respectively). Furthermore, photostimulations at 5–30 Hz were less effective at reducing SWS sleep-to-wake latencies in *Hcrt::ChR2-mCherry*-transduced *Hcrt* knockout mice compared with wild-type mice (Fig. 4c), demonstrating that Hcrt release is involved in the behavioural effects. However, the effect of photostimulation on wake latency seemed not to be completely blocked in *Hcrt* knockout animals, suggesting a possible role of other neurotransmitters (glutamate, dynorphin) co-expressed in Hcrt neurons.

This study directly establishes a causal link between electrical activity of Hcrt neurons and sleep-to-wake transitions. Using optogenetic technology, we found that we were able to target specifically Hcrt neurons for photostimulation *in vitro*, and that activation of Hcrt neurons at 5 Hz and above (but not at 1 Hz) resulted in a robust decrease in sleep-to-wake latencies from SWS or REM sleep *in vivo*.

Increased probability and number of transitions during repeated stimulation of Hcrt cells suggests a modulatory mechanism by which the Hcrt neural network promotes the initiation of arousal. Hcrt neurons might have additional roles under other conditions than those explored here in many aspects of sleep, including wake maintenance and other arousal-related behaviours. Arousal stability might result from Hcrt activation and divergent modulation of other arousal circuits such as the glutamatergic, cholinergic, histaminergic and dopaminergic systems. The effects seen with even unilateral stimulation suggest that Hcrt neurons have a potent and causal role in driving awakening; bilateral photostimulation could have an even stronger effect on the shortening of sleep-to-wake transitions or arousal maintenance. Dissection of the causal interactions of the

Hcrt system with these other neuromodulatory systems may now be possible with selective control of the Hcrt population. Because Hcrt deficiency results in arousal instability associated with narcolepsy, further insights into sleep disorders may result. Indeed, this demonstration that selective activation of genetically identified neurons can influence a complex behaviour like sleep raises the possibility that this optical technology could be extended to probing the circuit bases of other neurobehavioural and neuropsychiatric disorders.

## METHODS SUMMARY

**Plasmid and virus preparation.** The 3,086-base-pair (*EcoRI*–*SacI*) mouse *Hcrt* promoter<sup>19,20</sup> was used to replace the *CaMKII $\alpha$*  promoter in the *CaMKII $\alpha$ ::ChR2-mCherry* lentivirus vector<sup>25</sup>. The *Hcrt::mCherry* control viruses were made by swapping *ChR2-mCherry* with *mCherry* alone. High-titre lentiviruses were produced as described<sup>25</sup>.

**Slice preparation and electrophysiology.** Recombinant *Hcrt::ChR2-mCherry* lentivirus was stereotactically injected into the lateral hypothalamus (anteroposterior, 1.6 mm; mesolateral, 0.75 mm; dorsoventral, 5.0 mm)<sup>26</sup> of 5–6-week-old *Hcrt::EGFP* mice<sup>21</sup> as described<sup>27</sup>. Coronal slices (250  $\mu$ m) were prepared and Hcrt neurons were identified via GFP fluorescence and recorded as described<sup>21</sup>.

***In vivo* light delivery and sleep recordings.** All of the experiments described here were conducted during the same circadian period (12:00–18:00). Each stimulation was applied 15 s after the occurrence of a stable SWS or REM sleep event as detected by online EEG/EMG analysis. Latencies of light-induced sleep-to-wake transitions were measured by off-line scoring of the EEG/EMG recordings as described<sup>28</sup> (see Methods).

**Statistical methods.** Latencies of sleep-to-wake transitions were analysed using the Student's *t*-test. Pharmacological studies were analysed using a two-tailed paired Student's *t*-test. Two-way ANOVA was used to detect significance in the long-term stimulation experiments.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 2 February; accepted 1 October 2007.

Published online 17 October 2007.

1. Pace-Schott, E. F. & Hobson, J. A. The neurobiology of sleep: genetics, cellular physiology and subcortical networks. *Nature Rev. Neurosci.* **3**, 591–605 (2002).
2. Saper, C. B., Chou, T. C. & Scammell, T. E. The sleep switch: hypothalamic control of sleep and wakefulness. *Trends Neurosci.* **24**, 726–731 (2001).

3. de Lecea, L. *et al.* The hypocretins: hypothalamus-specific peptides with neuroexcitatory activity. *Proc. Natl Acad. Sci. USA* **95**, 322–327 (1998).
4. Sakurai, T. *et al.* Orexins and orexin receptors: a family of hypothalamic neuropeptides and G protein-coupled receptors that regulate feeding behavior. *Cell* **92**, 573–585 (1998).
5. Peyron, C. *et al.* Neurons containing hypocretin (orexin) project to multiple neuronal systems. *J. Neurosci.* **18**, 9996–10015 (1998).
6. Chemelli, R. M. *et al.* Narcolepsy in orexin knockout mice: molecular genetics of sleep regulation. *Cell* **98**, 437–451 (1999).
7. Lin, L. *et al.* The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene. *Cell* **98**, 365–376 (1999).
8. Peyron, C. *et al.* A mutation in a case of early onset narcolepsy and a generalized absence of hypocretin peptides in human narcoleptic brains. *Nature Med.* **6**, 991–997 (2000).
9. Thannickal, T. C. *et al.* Reduced number of hypocretin neurons in human narcolepsy. *Neuron* **27**, 469–474 (2000).
10. Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nature Neurosci.* **8**, 1263–1268 (2005).
11. Nagel, G. *et al.* Light activation of channelrhodopsin-2 in excitable cells of *Caenorhabditis elegans* triggers rapid behavioral responses. *Curr. Biol.* **15**, 2279–2284 (2005).
12. Zhang, F., Wang, L. P., Boyden, E. S. & Deisseroth, K. Channelrhodopsin-2 and optical control of excitable cells. *Nature Methods* **3**, 785–792 (2006).
13. Lima, S. Q. & Miesenböck, G. Remote control of behavior through genetically targeted photostimulation of neurons. *Cell* **121**, 141–152 (2005).
14. Aravanis, A. M. *et al.* An optical neural interface: *In vivo* control of rodent motor cortex with integrated fiberoptic and optogenetic technology. *J. Neural Eng.* **4**, S143–S156 (2007).
15. Petreanu, L., Huber, D., Sobczyk, A. & Svoboda, K. Channelrhodopsin-2-assisted circuit mapping of long-range callosal projections. *Nature Neurosci.* **10**, 663–668 (2007).
16. Arenkiel, B. R. *et al.* *In vivo* light-induced activation of neural circuitry in transgenic mice expressing channelrhodopsin-2. *Neuron* **54**, 205–218 (2007).
17. Wang, H. *et al.* High-speed mapping of synaptic connectivity using photostimulation in Channelrhodopsin-2 transgenic mice. *Proc. Natl Acad. Sci. USA* **104**, 8143–8148 (2007).
18. Bi, A. *et al.* Ectopic expression of a microbial-type rhodopsin restores visual responses in mice with photoreceptor degeneration. *Neuron* **50**, 23–33 (2006).
19. Conti, B. *et al.* Transgenic mice with a reduced core body temperature have an increased life span. *Science* **314**, 825–828 (2006).
20. Sakurai, T. *et al.* Structure and function of human prepro-orexin gene. *J. Biol. Chem.* **274**, 17771–17776 (1999).
21. Li, Y., Gao, X. B., Sakurai, T. & van den Pol, A. N. Hypocretin/Orexin excites hypocretin neurons via a local glutamate neuron-A potential mechanism for orchestrating the hypothalamic arousal system. *Neuron* **36**, 1169–1181 (2002).
22. Lee, M. G., Hassani, O. K. & Jones, B. E. Discharge of identified orexin/hypocretin neurons across the sleep–waking cycle. *J. Neurosci.* **25**, 6716–6720 (2005).
23. Mileykovskiy, B. Y., Kiyashchenko, L. I. & Siegel, J. M. Behavioral correlates of activity in identified hypocretin/orexin neurons. *Neuron* **46**, 787–798 (2005).
24. Smart, D. *et al.* SB-334867-A: the first selective orexin-1 receptor antagonist. *Br. J. Pharmacol.* **132**, 1179–1182 (2001).
25. Zhang, F. *et al.* Multimodal fast optical interrogation of neural circuitry. *Nature* **446**, 633–639 (2007).
26. Paxinos, G. & Franklin, K. *The Mouse Brain in Stereotaxic Coordinates* 2nd edn (Academic, New York, 2001).
27. Cetin, A., Komai, S., Eliava, M., Seeburg, P. H. & Osten, P. Stereotaxic gene delivery in the rodent brain. *Nature Protocols* **1**, 3166–3173 (2006).
28. Bourgin, P. *et al.* Hypocretin-1 modulates rapid eye movement sleep through activation of locus coeruleus neurons. *J. Neurosci.* **20**, 7760–7765 (2000).
29. Eggemann, E. *et al.* The wake-promoting hypocretin-orexin neurons are in an intrinsic state of membrane depolarization. *J. Neurosci.* **23**, 1557–1562 (2003).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank S. Nishino and N. Fujiki for their technical support in sleep recording (Stanford University SCORE facility), and Y. Xu and C. E. Olin for critical comments. We also thank T. Sakurai and M. Yanagisawa for providing the *Hcr::EGFP* transgenic and *Hcr* knockout mice. A.R.A. is supported by the Belgian American Educational Foundation and the Fondation Leon Fredericq. F.Z. is supported by a fellowship from the NIH. A.M.A. is supported by the Walter and Idun Berry Foundation. K.D. is supported by NARSAD, APIRE and the Snyder, Culpeper, Coulter, Klingenstein, Whitehall, McKnight and Albert Yu and Mary Bechmann Foundations, as well as by NIMH, NIDA and the NIH Director's Pioneer Award Program. L.d.L. is supported by NIMH and NIDA.

**Author Contributions** All authors designed the experiments. A.R.A., F.Z. and A.M.A. collected data and performed analysis. All authors discussed the results and contributed to the text.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to L.d.L. (llecea@stanford.edu) or K.D. (deissero@stanford.edu).



## METHODS

**Animals.** Mice were housed in individual plexiglass recording cages in temperature- ( $22 \pm 1^\circ\text{C}$ ) and humidity (40–60%)–controlled recording chambers (custom-designed stainless steel cabinets with individual ventilated compartments) under a 12 h/12 h light/dark cycle (starting at 7:00). Food and water were available *ad libitum*. All the experiments were carried out in accordance with the guidelines described in the National Institutes of Health Guide for the Care and Use of Laboratory Animals.

**Immunocytochemistry.** To verify the phenotype of neurons that expressed the ChR2–mCherry fusion protein, infused transgenic *Hcrt::EGFP* mice were anaesthetized with isoflurane and perfused transcardially with physiological saline followed by 4% paraformaldehyde in PBS (pH 7.4). The brains were dissected out and post-fixed in the same fixative overnight at  $4^\circ\text{C}$  and finally cryoprotected in 30% sucrose dissolved in PBS for 48 h at  $4^\circ\text{C}$ . Thirty-micrometre coronal sections containing the lateral hypothalamus were washed in phosphate-buffered saline (PBS; pH 7.4), treated with 0.3% Triton X-100 (PBST), and immersed in a blocking solution consisting of 4% bovine serum albumin dissolved in PBS. Some sections were immunostained with rabbit antiserum against DsRed proteins (1:1,000; Clontech). A secondary anti-rabbit immunoglobulin conjugated to biotin was used before an amplification of the signal using the ABC kit (Vector) and TSA kit (NEN Life Science Products). Briefly, sections were washed in PBST for  $3 \times 10$  min, and incubated for 1 h at room temperature with an anti-rabbit biotinylated conjugate diluted in PBST. Sections were then washed three times in PBST and incubated with the TSA kit for 15 min. After washes, sections were finally incubated with a streptavidin–Alexa-fluor 596 conjugate (Molecular Probes) diluted 1:1,000 in PBST for 1 h at room temperature.

c-Fos and Hcrt double immunostaining was performed on brain sections from *Hcrt::ChR2* and *Hcrt::mCherry* transduced animals ( $n = 5$  and  $6$ , respectively). Briefly, 2–3 weeks after lentivirus injection, animals were photostimulated (light pulse trains of 10 s at 20 Hz every minute for 10 min) and perfused 1 h after stimuli. Brains were post-fixed and cryoprotected as described above. Coronal brain sections containing the lateral hypothalamus were successively incubated in (1) a rabbit antiserum to c-Fos (1:1,000, Calbiochem) in PBST supplemented with 4% bovine serum albumin (BSA, Sigma) for 1 day at  $4^\circ\text{C}$ ; (2) a biotinylated anti-rabbit IgG solution (1:1,000, Vector Laboratories) in PBST; and (3) an ABC-peroxidase solution (1:1,000, Vector Laboratories) both for 60 min at room temperature. Finally, sections were stained using 3,3'-diaminobenzidine-4 HCl and nickel solution (DAB-Ni; Vector Laboratories) to obtain dark-purple staining. Four washes in phosphate buffer were performed between each step. The c-Fos-stained sections were incubated in a goat antiserum to Hcrt (1:5,000; Santa Cruz) in PBST/BSA 4% for 2 days at  $4^\circ\text{C}$ . Amplification steps were similar to those described above but the final step was performed in DAB solution without nickel to obtain brown staining. Finally, the sections were mounted on slides, dried and coverslipped with permaslip.

Sections were washed, mounted and coverslipped, and examined with a fluorescence microscope (Carl Zeiss). Digital images from the microscopy were slightly modified to optimize for image resolution, brightness and contrast in Adobe Photoshop 7.0 (Adobe Systems), to represent optimally the immunohistochemistry observed.

Hcrt-positive and Hcrt-positive/c-Fos-positive neurons were counted in each hypothalamic structure (within both ipsilateral and contralateral sides of injection site) present on at least four Hcrt/c-Fos double-labelled sections for each animal. Lateral hypothalamus coronal section maps were made according to the

mouse brain atlas<sup>26</sup> to represent the placement of the cannula guide in each animal.

**Polysomnographic recording and analysis.** For sleep–wake cycle recording, the EEG signals were recorded from two electrodes placed on the frontal (AP,  $-2$ ; ML,  $\pm 1$ ) and temporal (AP,  $3$ ; ML,  $\pm 2.5$ ) cortices. Two wire electrodes were inserted in the neck musculature to record postural tone through EMG activity. Insulated leads from the EEG and EMG electrodes were then soldered to a miniconnector that was cemented to the skull with meta-bond (Parkell) and dental acrylic. Mice were housed individually for at least 10 days after surgery and were then connected to commutators with flexible cables for habituation to the EEG/EMG recording conditions. Cortical EEG and EMG signals were amplified using Grass Instruments (West Warwick) and digitized at 256 Hz using a sleep recording system (VitalRecorder, Kissei Comtec). The signals were digitally filtered and spectrally analysed by fast Fourier transform (represented in Fig. 3a, right panels), and polysomnographic recordings were visually scored by 5-s epochs for wake, SWS and REM sleep, as previously described<sup>28</sup>. The following criteria were used. Wakefulness is characterized by desynchronized low-amplitude EEG and EMG activity with phasic bursts. SWS is characterized by synchronized, high-amplitude, low-frequency (0.25–4 Hz) EEG and reduced EMG activity compared to wakefulness. A decrease in the EEG amplitude associated with a flat EMG (muscle atonia) and a regular and pronounced theta rhythm (4–9 Hz) signalled the onset of REM sleep episodes. Sleep state changes were recorded when at least one epoch in a different sleep stage was scored. SWS and REM sleep-to-wake transition latencies were calculated as the time from the end of the light stimulation (during SWS or REM sleep) to the first event of wakefulness. Awakenings occurring during photostimulation were discarded from the analysis. The EEG power densities obtained for each state were summed over the frequency band of 0.5–40 Hz (total power). To normalize the data, all power spectral densities at the different frequency ranges, that is, delta (0.5–4 Hz), theta (4.5–9 Hz), alpha (9–15 Hz), were expressed as relative values to the total power of the same epochs. Four-second epochs with more than two vigilance states were omitted from the EEG power analysis. The full sleep–wake cycle of mice is very short compared to human (2–3 min). Consequently, based on the EEG/EMG precise criteria and power spectrum analysis of the EEG, mice show sleep epochs between 5 and 250 s in duration, with an average of 50 s during spontaneous sleep.

Optical fibres were placed 1 h before experiments (11:00) and removed immediately afterwards (18:00). Light pulse trains (1–30 Hz) were programmed using a function generator (33220A, Agilent Technologies). For each animal, spontaneous sleep–wake cycles as well as light-induced sleep-to-wake transitions were characterized. In a first set of experiments, we measured latencies of light-induced sleep-to-wake transitions at different frequencies (1–30 Hz and 10 s continuous illumination). In a second series of experiments, specificity of the activation of Hcrt neural network on the light-induced SWS and REM sleep-to-wake transitions was assessed by administration of the Hcrt receptor 1 antagonist (SB334867, Tocris) before photostimulation (15 ms light pulses at 20 Hz during 10 s). Latency to wakefulness was measured as previously described. In a third series of experiments, we used an automated long-term stimulation protocol (15 ms light pulses at 20 Hz for 10 s every minute during 1 h) to test the behavioural consequences of sustained stimulations of the Hcrt neural network. The sleep–wake cycle parameters (wake, SWS, REM sleep duration and behavioural transitions) were then quantified by off-line scoring of the entire hour of stimulation and the corresponding circadian baseline for each animal. Polysomnographic scoring was performed blindly by two scorers and was found to lie within a 95% confidence interval ( $\kappa$  coefficient = 0.840).

# A synaptic memory trace for cortical receptive field plasticity

Robert C. Froemke<sup>1</sup>, Michael M. Merzenich<sup>1</sup> & Christoph E. Schreiner<sup>1</sup>

Receptive fields of sensory cortical neurons are plastic, changing in response to alterations of neural activity or sensory experience<sup>1–12</sup>. In this way, cortical representations of the sensory environment can incorporate new information about the world, depending on the relevance or value of particular stimuli<sup>1,6,9</sup>. Neuromodulation is required for cortical plasticity, but it is uncertain how subcortical neuromodulatory systems, such as the cholinergic nucleus basalis, interact with and refine cortical circuits<sup>13–24</sup>. Here we determine the dynamics of synaptic receptive field plasticity in the adult primary auditory cortex (also known as AI) using *in vivo* whole-cell recording. Pairing sensory stimulation with nucleus basalis activation shifted the preferred stimuli of cortical neurons by inducing a rapid reduction of synaptic inhibition within seconds, which was followed by a large increase in excitation, both specific to the paired stimulus. Although nucleus basalis was stimulated only for a few minutes, reorganization of synaptic tuning curves progressed for hours thereafter: inhibition slowly increased in an activity-dependent manner to rebalance the persistent enhancement of excitation, leading to a retuned receptive field with new preference for the paired stimulus. This restricted period of disinhibition may be a fundamental mechanism for receptive field plasticity, and could serve as a memory trace<sup>9,25</sup> for stimuli or episodes that have acquired new behavioural significance.

A major subcortical nucleus critical for receptive field plasticity is nucleus basalis, the main source of cortical acetylcholine<sup>4,9,14–17,19–21</sup>. How are neuromodulators such as acetylcholine involved in plasticity, and what circuit elements do they act on? One possibility is that neuromodulation creates a cellular tag or memory trace for synaptic events that occurred in conjunction with neuromodulator release. However, the effects of acetylcholine on cortical neurons are diverse, including increased excitability<sup>19,22,23</sup> and suppression of synaptic transmission<sup>15,17,23,24</sup>, and it is unclear how these effects could produce long-term response enhancement specific for particular stimuli. Extracellular recordings cannot reveal which cellular events are responsible for receptive field plasticity, and studies *in vitro* do not permit investigation of receptive fields or subcortical interactions with cortical networks. Instead, here we use whole-cell recording and nucleus basalis stimulation in the intact brain to determine the synaptic basis of cortical receptive field plasticity.

We made *in vivo* whole-cell recordings from adult rat primary auditory cortex (Fig. 1a). Pure tones of different frequencies were played in pseudo-random sequence to the contralateral ear. Frequency tuning was characterized in voltage-clamp at hyperpolarized (–70 mV) and depolarized (–20 mV) levels to reveal tone-evoked excitatory and inhibitory postsynaptic currents (EPSCs and IPSCs), respectively (Fig. 1b).

Initially, cortical neurons had similar profiles of excitatory and inhibitory frequency tuning, confirming that their levels of

excitation and inhibition were balanced<sup>26,27</sup>. Excitatory and inhibitory tuning curves usually had one shared peak at the best frequency. There was a high correlation between the relative amounts of excitation and inhibition across all frequencies (Supplementary Fig. 1).

After 5–15 min of baseline receptive field characterization, a tone was paired repetitively for 2–5 min with electrical stimulation of nucleus basalis ('nucleus basalis pairing', Fig. 1b) to release endogenous acetylcholine within primary auditory cortex (although other substances may also be released<sup>28</sup>). Tetanic stimulation (100 Hz, 250 ms) was performed in a manner similar to natural spiking patterns of nucleus basalis cholinergic cells<sup>20</sup>.

After cessation of nucleus basalis pairing, we observed large changes to synaptic inputs evoked by the paired tone: pairing rapidly potentiated tone-evoked EPSCs and depressed IPSCs (Fig. 1c, d). Similar results were obtained when conductance and charge transfer were measured (Supplementary Figs 2 and 3). Synaptic modifications were long lasting and frequency specific (but see below). On average, currents evoked by unpaired tones were not significantly altered, although we consistently observed that responses to the original best frequency were reduced over a longer time course (Fig. 1d). Thus the main effect of nucleus basalis pairing is to break the balance between excitation and inhibition at the paired frequency.

Synaptic modifications required paired nucleus basalis and sensory stimulation. Frequency tuning was not persistently altered when nucleus basalis was stimulated in silence or when a given tone was repeated without nucleus basalis stimulation (Supplementary Fig. 4a, b, d). Surprisingly, this was the case even in current-clamp recordings in which tone presentation reliably evoked postsynaptic spikes (Supplementary Fig. 4c, d), demonstrating that repetitive pairing of tones with postsynaptic spikes does not induce long-term potentiation of EPSPs at these synapses.

How do the synaptic modifications observed here correspond to previously reported changes in excitability<sup>17,22</sup> and spike generation<sup>4,9,19</sup>? To determine the relation between changes in synaptic input and spike output, we made current-clamp recordings from primary auditory cortex neurons to measure tone-evoked spiking responses before and after nucleus basalis pairing. Before pairing, tones generally evoked subthreshold EPSPs or a single spike<sup>26,27,29</sup>. As expected, pairing increased spiking evoked by the paired frequency. This included a dramatic (>7-fold) increase in the probability of firing bursts of 2+ spikes (Supplementary Fig. 5). Consistent with the specific increase in firing rate, pairing had no long-term effect on input resistance ( $R_i$ ; Supplementary Fig. 5b). Thus, nucleus basalis pairing alters the firing mode of cortical neurons, increasing the output of primary auditory cortex to enhance the salience of particular stimuli such as those with new behavioural relevance<sup>4,9,18</sup>, or during periods of increased perceptual demand on attention<sup>15,30</sup>.

<sup>1</sup>Coleman Memorial Laboratory and W. M. Keck Foundation Center for Integrative Neuroscience, Department of Otolaryngology, University of California, San Francisco, California 94143, USA.

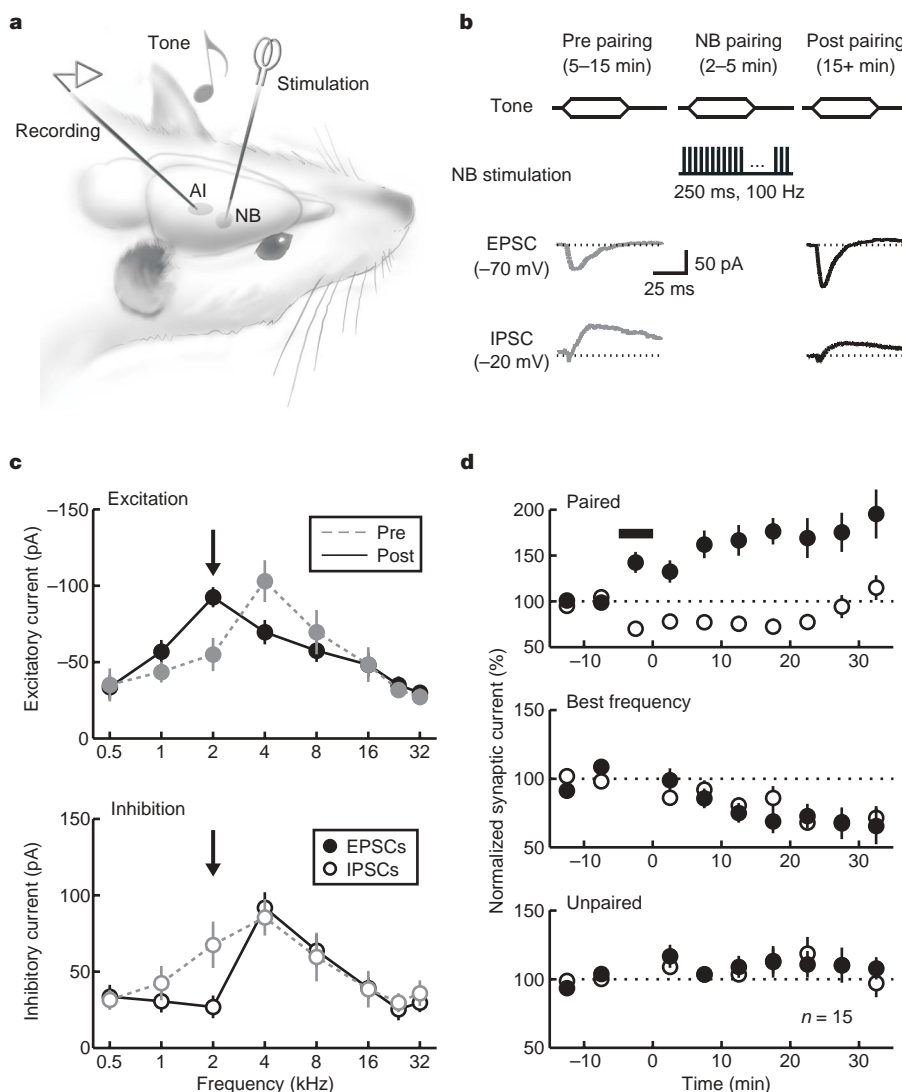
Synaptic modifications and enhancement of spiking occurred not only after nucleus basalis pairing, but also during pairing itself. To determine the time course of pairing-induced changes, we examined the responses to the paired tone during the pairing procedure (Fig. 2). Nucleus basalis pairing suppressed IPSCs within twenty seconds, but enhancement of EPSCs took approximately three times as long (Fig. 2a–c). Cortical application of atropine, an acetylcholine receptor antagonist, prevented the effects of nucleus basalis pairing (Fig. 2d)<sup>9,17</sup>. Therefore, despite the existence of multiple transmitter systems in nucleus basalis<sup>28</sup> and the heterogeneity of cholinergic neuromodulation<sup>15,17,19,22–24</sup>, the net effects of nucleus basalis pairing are suppression of inhibition followed by enhancement of excitation. These results suggest that a central role for nucleus basalis activation in receptive field plasticity is to trigger spectrotemporally restricted disinhibition, permissive for induction of Hebbian synaptic plasticity<sup>1,8,9,11,17</sup>.

We wondered which inputs were modified after nucleus basalis pairing. The decoupling of inhibition from excitation implied that a primary site of synaptic modification was directly within primary auditory cortex. However, it is unclear to what degree extrinsic or

intrinsic projections mediate cortical plasticity<sup>7,9,11,15,17,24</sup>. To localize the effects of nucleus basalis pairing, we used an additional pair of stimulation electrodes to concurrently monitor two distinct inputs: one from the ventral division of the thalamic medial geniculate body, and one from primary auditory cortex (Fig. 3a). We ensured that both stimulation electrodes were in co-tuned areas by making extracellular recordings of receptive fields through the electrodes.

We initially recorded electrically evoked synaptic currents by intracortical and thalamic stimulation for 5–10 min in the absence of sensory stimulation. Then, electrical stimulation was stopped, and nucleus basalis stimulation was paired with the best frequency at the sites of thalamic and intracortical stimulation for 2–5 min. Finally, sensory and nucleus basalis stimulation were stopped, and electrical stimulation was resumed (Fig. 3b).

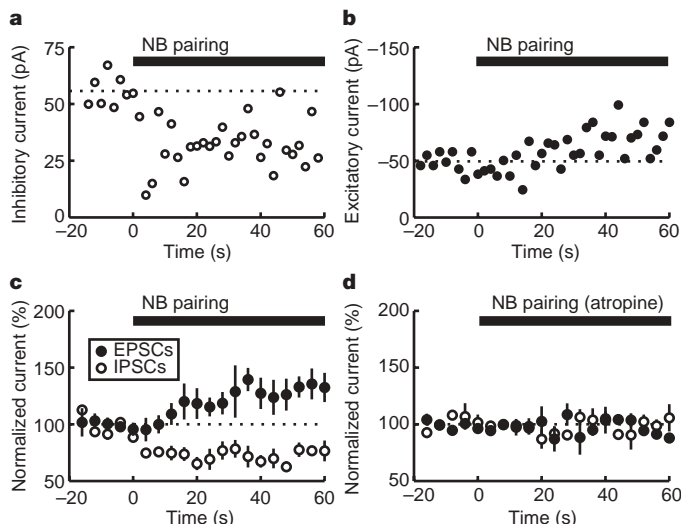
Nucleus basalis pairing persistently modified synaptic currents evoked by intracortical stimulation (Fig. 3c, e) but not thalamic stimulation (Fig. 3d, f). These modifications were similar in sign, magnitude and duration to changes in tone-evoked synaptic responses. Intracortical EPSCs were potentiated, whereas intracortical IPSCs were suppressed (Fig. 3e). Thalamocortical EPSCs were



**Figure 1 | Synaptic modifications induced by nucleus basalis pairing.** **a**, Experimental configuration. **b**, Experimental procedure. **c**, Example of pairing-induced modification of synaptic tuning curves. Upper panel, excitatory tuning. EPSCs at the paired frequency (2 kHz) increased from  $-54.9 \pm 10.9$  to  $-92.4 \pm 6.6$  pA (68.4%,  $P < 0.006$ ,  $t$ -test; filled symbols). Lines, tuning before (dashed grey) and ~10 min after (solid black) pairing. Arrow, paired tone. Lower panel, inhibitory tuning. IPSCs at the paired

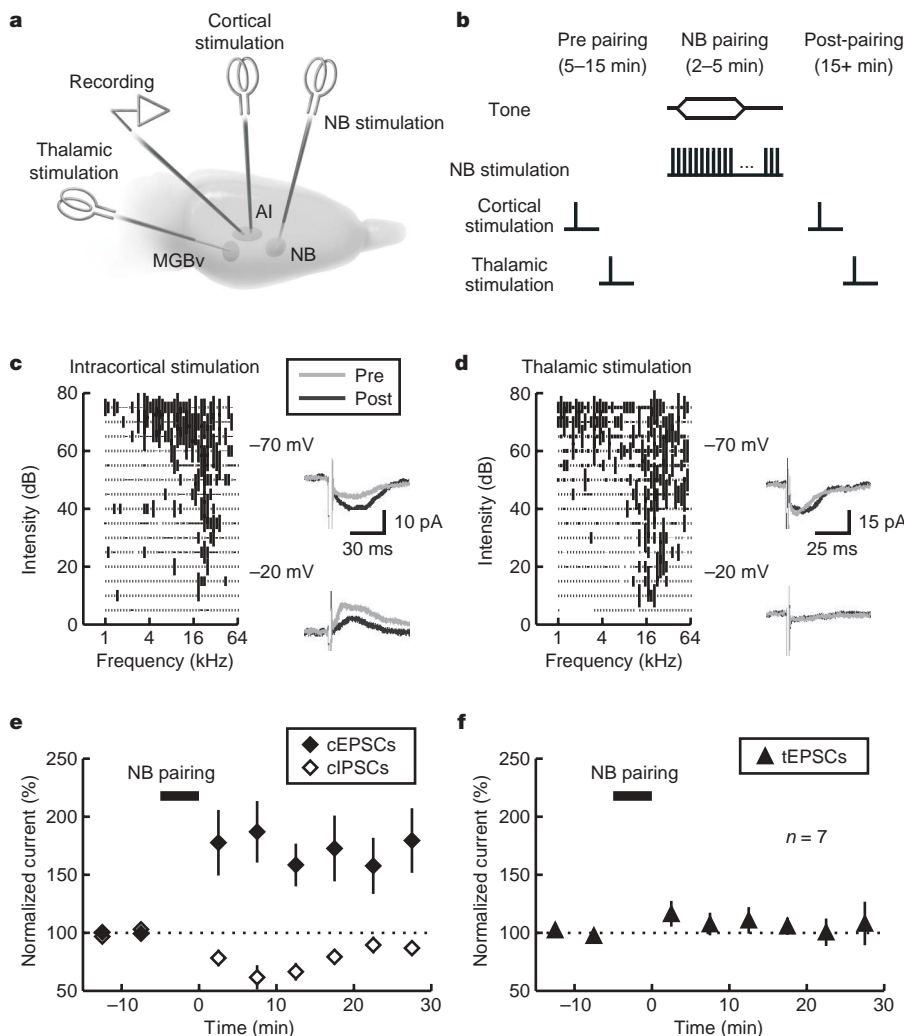
frequency decreased from  $67.6 \pm 15.2$  to  $27.0 \pm 7.4$  pA (–60.1%,  $P < 0.03$ ; open symbols). **d**, Time course. Uppermost panel, paired frequency (excitation:  $68.0 \pm 13.9\%$ ,  $n = 15$ ,  $P < 0.0007$ ; inhibition:  $-24.8 \pm 6.0\%$ ,  $P < 0.0002$ ). Horizontal bar, nucleus basalis pairing. Middle panel, original best frequency. Lowest panel, other unpaired tones. Error bars, s.e.m. NB, nucleus basalis; AI, primary auditory cortex.





**Figure 2 | Rapid suppression of inhibition during nucleus basalis pairing.**

**a**, IPSCs decreased from  $55.7 \pm 2$  to  $32.3 \pm 4$  pA ( $-41.9\%$ ,  $P < 0.002$ ) 41–60 s after pairing. Solid line, nucleus basalis pairing. Dashed line, mean current before pairing. **b**, EPSCs increased from  $-49.5 \pm 3$  to  $-70.3 \pm 5$  pA ( $42.0\%$ ,  $P < 0.01$ ). **c**, Suppression of inhibition occurred before enhancement of excitation (inhibition,  $20.7 \pm 9.3$  s; excitation,  $52.4 \pm 4.1$  s,  $P < 0.009$ ). Filled symbols, excitation (1–20 s,  $6.5 \pm 8.6\%$ ,  $n = 9$ ,  $P > 0.4$ ; 41–60 s,  $29.8 \pm 2.2\%$ ,  $P < 0.003$ ); open symbols, inhibition (1–20 s,  $-24.5 \pm 3.8\%$ ,  $n = 6$ ,  $P < 0.005$ ; 41–60 s,  $-28.1 \pm 7.0\%$ ,  $P < 0.02$ ). **d**, Atropine (1 mM) blocked the effects of nucleus basalis pairing ( $n = 4$ ,  $P > 0.5$ ). Error bars, s.e.m.



**Figure 3 | Nucleus basalis pairing altered intracortical connections.** **a**, Experimental configuration. MGBv, ventral division of the thalamic medial geniculate body. **b**, Experimental procedure. **c**, Intracortical stimulation. Left panel, receptive field recorded with stimulation electrode (16 kHz region of AI). Right panel, whole-cell recording from the 6 kHz region of AI, before (grey) and after (black) pairing nucleus basalis stimulation with 16 kHz tones. Intracortical EPSCs (cEPSCs) increased ( $73.4\%$ ;  $P < 0.03$ ) and intracortical IPSCs (cIPSCs) decreased ( $-47.1\%$ ;  $P < 0.03$ ). **d**, Same experiment as **c**, but thalamic stimulation. Thalamocortical EPSCs (tEPSCs) were unchanged ( $P > 0.5$ ). **e**, Nucleus basalis pairing enhanced cEPSCs ( $72.8 \pm 20.5\%$ ,  $n = 7$ ,  $P < 0.02$ ; filled) and suppressed cIPSCs ( $-30.7 \pm 3.6\%$ ,  $P < 0.004$ ; open). **f**, Same experiments as **e**, but for tEPSCs ( $P > 0.4$ ). Error bars, s.e.m.

unaffected (Fig. 3f), and thalamocortical IPSCs were not observed. These data suggest that nucleus basalis pairing does not induce strengthening of direct thalamocortical input from neurons of the ventral division of the thalamic medial geniculate body that are tuned to the paired frequency, but it does enhance connections from the region of primary auditory cortex initially tuned to that frequency. The decrease in intracortical IPSC amplitude demonstrates that one location of synaptic modification is directly within primary auditory cortex, at the connections between interneurons and excitatory cells. However, these results do not exclude potential for modification of other synapses elsewhere in the auditory pathway, perhaps on a different timescale or with other requirements for induction<sup>7,9</sup>.

Finally, we noticed that towards the end of long-term recording sessions ( $\sim 30$  min after pairing; Fig. 1d), IPSCs evoked by tones of the paired frequency seemed to recover back towards their initial amplitudes. This indicated that modification of inhibitory frequency tuning occurred with more complex dynamics than enhancement of excitation. However, as it was difficult to maintain stable recordings for longer than 30+ minutes, we were unable to follow the complete evolution of inhibitory modifications within individual recordings.

To examine the time course of synaptic receptive field plasticity thoroughly, we made consecutive whole-cell recordings from the same location in primary auditory cortex for hours after nucleus basalis pairing in each animal (Fig. 4). To compare synaptic modifications across cells, we took advantage of the consistency of frequency tuning for neurons in a given tonotopic region of primary auditory cortex (Supplementary Fig. 1b), and normalized current amplitudes to their maximum values across frequencies.

For example, the recordings shown in Fig. 4a–d were each made from the 16 kHz region of primary auditory cortex in the same animal. As expected, best frequencies of excitation and inhibition for the first recorded neuron were both initially 16 kHz (Fig. 4a, open arrowhead). After pairing nucleus basalis stimulation with 4 kHz tones, we observed a large increase in the excitation–inhibition ratio (E:I ratio) at the paired frequency (Fig. 4b, arrow). After this recording was finished, we obtained another recording from a second cell in the same location ~100 min after pairing (Fig. 4c). The best frequency of excitation for this cell was 4 kHz, but inhibition was maximal at 8 kHz. Finally, we recorded a third cell, in the same location as the two previous recordings, ~3 h after pairing (Fig. 4d), and found that best frequencies of excitation and inhibition were 4 kHz. Thus potentiation of excitation was maintained for hours after transient nucleus basalis pairing, but after an initial period of suppression, inhibition began to progressively increase until it balanced the enhanced excitation.

Re-establishing a normal E:I ratio required approximately two hours after completion of nucleus basalis pairing (Fig. 4e, squares). This rebalancing reflects the gradual growth of inhibitory strength rather than a decrease in excitation at the paired frequency, was apparent for continuously recorded neurons (Fig. 4e, circles), and was registered as shifts in best frequency (Supplementary Fig. 6). Rebalancing required near-continual tonal stimulation. If auditory stimulation was turned off for 60–90 min following nucleus basalis pairing, suppression of inhibition was maintained and the E:I ratio remained unbalanced (Fig. 4f, ‘Quiet’). These data are reminiscent of

recent findings showing that the timing of the primary auditory cortex critical period can be altered by exposure to different auditory environments<sup>10,12</sup>.

We have described here a differential progression for changes in cortical excitation and inhibition after nucleus basalis pairing that reorganizes primary auditory cortex receptive fields. Although nucleus basalis was stimulated only for a brief period, alteration of excitatory frequency tuning required 30+ minutes to manifest fully, leading to increased preference for paired stimuli. Changes to inhibitory tuning, however, occurred first and continued for hours after nucleus basalis pairing, eventually increasing to balance the changes in excitation. These results provide a mechanism for the function of nucleus basalis in attentional modulation: focal disinhibition may act as a synaptic correlate of heightened attentiveness for novel or meaningful stimuli. Furthermore, the long-lasting break in the E:I balance caused by nucleus basalis pairing maintains the immediate effects of nucleus basalis activation, allowing restricted parts of cortex to operate in hyperexcitable states independent of further neuromodulation. This transient disinhibition therefore acts as a synaptic memory trace for sensory information of increased significance<sup>9,25,30</sup>, allowing these stimuli to evoke larger bursts of spikes for a limited time while receptive fields are adjusted to represent the new emphasis for paired inputs.

## METHODS SUMMARY

Experimental procedures were approved under UCSF IACUC protocols. Experiments were carried out in a sound-attenuating chamber. Female Sprague-Dawley rats 3–5 months old were anaesthetized with pentobarbital. A stimulation electrode was implanted in right nucleus basalis<sup>4</sup> and right auditory cortex was exposed. Pure tones (0.5–32 kHz, 50 ms duration, 60–80 dB) in pseudo-random sequence were delivered into the left ear canal by a tube sealed to a calibrated speaker. The location of primary auditory cortex was determined by mapping spike responses using tungsten electrodes<sup>4,12</sup>.

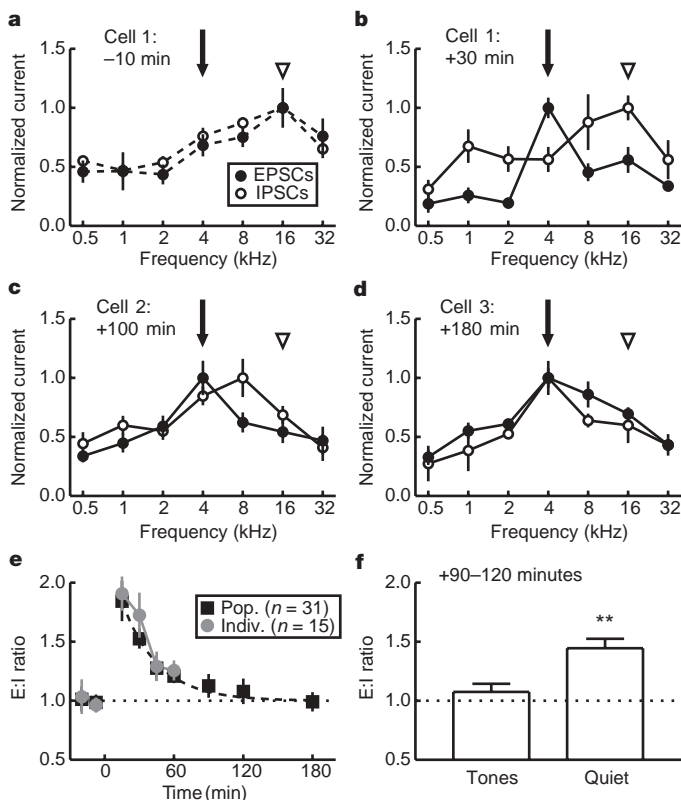
*In vivo* whole-cell recordings were obtained from primary auditory cortex neurons located 400–1,100  $\mu$ m below the pial surface. Patch pipettes (5–9 M $\Omega$ ) contained (in mM): 125 Cs-gluconate, 5 TEACl, 4 MgATP, 0.3 GTP, 10 phosphocreatine, 10 HEPES, 0.5 EGTA, 3.5 QX-314, 2 CsCl, pH 7.2 (voltage-clamp); or: 135 K-gluconate, 5 NaCl, 5 MgATP, 0.3 GTP, 10 phosphocreatine, 10 HEPES, 0.5 EGTA, pH 7.3 (current-clamp). Resting potential,  $-66.0 \pm 10$  mV (s.d.);  $R_i$ ,  $105.1 \pm 54$  M $\Omega$ .

To make consecutive recordings from the same location of primary auditory cortex, subsequent electrodes were positioned at the same penetrations. Currents were normalized to the largest across frequencies, and the E:I ratio ( $EPSC_{paired}/EPSC_{BF}$ )/( $IPSC_{paired}/IPSC_{BF}$ ) was calculated (BF, best frequency). Frequency tuning was sampled at 0.5–1 octave intervals, outside of the normal variance in best frequency for a given location in primary auditory cortex.

For microstimulation, stimulation strengths were set at the minimum required ( $\leq 20$   $\mu$ A) to reliably evoke small synaptic events. Intracortical stimulation electrodes were placed 400–1,000  $\mu$ m from recording electrodes. Thalamic stimulation electrodes were implanted in the right ventral division of the thalamic medial geniculate body.

Received 17 April; accepted 21 September 2007.

- Buonomano, D. V. & Merzenich, M. M. Cortical plasticity: from synapses to maps. *Annu. Rev. Neurosci.* **21**, 149–186 (1998).
- Debanne, D., Shulz, D. E. & Fregnac, Y. Activity-dependent regulation of ‘on’ and ‘off’ responses in cat visual cortical receptive fields. *J. Physiol. (Lond.)* **508**, 523–548 (1998).
- Gilbert, C. D. Adult cortical dynamics. *Physiol. Rev.* **78**, 467–485 (1998).
- Kilgard, M. P. & Merzenich, M. M. Cortical map reorganization enabled by nucleus basalis activity. *Science* **279**, 1714–1718 (1998).
- Chang, E. F. & Merzenich, M. M. Environmental noise retards auditory cortical development. *Science* **300**, 498–502 (2003).
- Fritz, J., Shamma, S., Elhilali, M. & Klein, D. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature Neurosci.* **6**, 1216–1223 (2003).
- Suga, N. & Ma, X. Multiparametric corticofugal modulation and plasticity in the auditory system. *Nature Rev. Neurosci.* **4**, 783–794 (2003).
- Malenka, R. C. & Bear, M. F. LTP and LTD: an embarrassment of riches. *Neuron* **44**, 5–21 (2004).
- Weinberger, N. M. Specific long-term memory traces in primary auditory cortex. *Nature Rev. Neurosci.* **5**, 279–290 (2004).



**Figure 4 | Temporal dynamics of synaptic receptive field plasticity.**

**a**, Frequency tuning before pairing. Arrow, paired frequency (4 kHz; E:I ratio, 0.9); arrowhead, original best frequency (16 kHz). **b**, Same cell as **a**, 30 min after pairing (E:I ratio, 1.78). **c**, Second cell, 100 min after pairing (E:I ratio, 1.18). **d**, Third cell, 180 min after pairing (E:I ratio: 1.0). **e**, Rebalance of E:I ratio at paired frequency. Squares, consecutive recordings from cell populations at the same locations (Pop.); dashed line, exponential fit; circles, individual continuous recordings from Fig. 1d (Indiv.). **f**, E:I ratio 90–120 min after pairing (Tones, E:I ratio increase of  $7.5 \pm 6.9\%$ ,  $n = 12$ ; Quiet, E:I ratio increase of  $44.6 \pm 7.9\%$ ,  $n = 6$ ,  $P < 0.004$ ). Double asterisks,  $P < 0.01$ . Error bars, s.e.m.

10. Chang, E. F., Bao, S., Imaizumi, K., Schreiner, C. E. & Merzenich, M. M. Development of spectral and temporal response selectivity in the auditory cortex. *Proc. Natl Acad. Sci. USA* **102**, 16460–16465 (2005).
11. Karmarkar, U. R. & Dan, Y. Experience-dependent plasticity in adult visual cortex. *Neuron* **52**, 577–585 (2006).
12. de Villers-Sidani, E., Chang, E. F., Bao, S. & Merzenich, M. M. Critical period window for spectral tuning defined in the primary auditory cortex (A1) in the rat. *J. Neurosci.* **27**, 180–189 (2007).
13. Bear, M. F. & Singer, W. Modulation of visual cortical plasticity by acetylcholine and noradrenaline. *Nature* **320**, 172–176 (1986).
14. Everitt, B. J. & Robbins, T. W. Central cholinergic systems and cognition. *Annu. Rev. Psychol.* **48**, 649–684 (1997).
15. Sarter, M., Hasselmo, M. E., Bruno, J. P. & Givens, B. Unraveling the attentional functions of cortical cholinergic inputs: interactions between signal-driven and cognitive modulation of signal detection. *Brain Res. Brain Res. Rev.* **48**, 98–111 (2005).
16. Zhang, Y., Hamilton, S. E., Nathanson, N. M. & Yan, J. Decreased input-specific plasticity of the auditory cortex in mice lacking M1 muscarinic acetylcholine receptors. *Cereb. Cortex* **16**, 1258–1265 (2006).
17. Rasmusson, D. D. The role of acetylcholine in cortical synaptic plasticity. *Behav. Brain Res.* **115**, 205–218 (2000).
18. Yu, A. J. & Dayan, P. Uncertainty, neuromodulation, and attention. *Neuron* **46**, 681–692 (2005).
19. Metherate, R., Cox, C. L. & Ashe, J. H. Cellular bases of neocortical activation: modulation of neural oscillations by the nucleus basalis and endogenous acetylcholine. *J. Neurosci.* **12**, 4701–4711 (1992).
20. Lee, M. G., Hassani, O. K., Alonso, A. & Jones, B. E. Cholinergic basal forebrain neurons burst with theta during waking and paradoxical sleep. *J. Neurosci.* **25**, 4365–4369 (2005).
21. Richardson, R. T. & DeLong, M. R. A reappraisal of the functions of the nucleus basalis of Meynert. *Trends Neurosci.* **11**, 264–267 (1988).
22. Woody, C. D. & Gruen, E. Acetylcholine reduces net outward currents measured *in vivo* with single electrode voltage clamp techniques in neurons of the motor cortex of cats. *Brain Res.* **424**, 193–198 (1987).
23. Xiang, Z., Huguenard, J. R. & Prince, D. A. Cholinergic switching within neocortical inhibitory networks. *Science* **281**, 985–988 (1998).
24. Metherate, R. *et al.* Spectral integration in auditory cortex: mechanisms and modulation. *Hear. Res.* **206**, 146–158 (2005).
25. Thompson, R. F. In search of memory traces. *Annu. Rev. Psychol.* **56**, 1–23 (2005).
26. Wehr, M. & Zador, A. M. Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature* **426**, 442–446 (2003).
27. Tan, A. Y. Y., Zhang, L. I., Merzenich, M. M. & Schreiner, C. E. Tone-evoked excitatory and inhibitory synaptic conductances of primary auditory cortex neurons. *J. Neurophysiol.* **92**, 630–643 (2004).
28. Gritti, I., Manns, I. D., Mainville, L. & Jones, B. E. Parvalbumin, calbindin, or calretinin in cortically projecting and GABAergic, cholinergic, or glutamatergic basal forebrain neurons of the rat. *J. Comp. Neurol.* **458**, 11–31 (2003).
29. DeWeese, M. R., Wehr, M. & Zador, A. M. Binary spiking in auditory cortex. *J. Neurosci.* **23**, 7940–7949 (2003).
30. Hromádka, T. & Zador, A. M. Towards the mechanisms of auditory attention. *Hear. Res.* **229**, 180–185 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank K. L. Arendt, T. Babcock, Y. Dan, E. de Villers-Sidani, M. R. DeWeese, M. P. Kilgard, D. Polley, L. Wilbrecht and J. A. Winer for comments and discussions, and S. Bao, W. Huang, K. Imaizumi, A. Tan and C.-L. Teng for technical assistance. D. Bliss created the artwork in Figs 1a and 3a. This work was supported by the NIDCD, the Conte Center for Neuroscience Research at UCSF, Hearing Research Inc., and the John C. and Edward Coleman Fund. R.C.F. is a recipient of the Jane Coffin Childs Postdoctoral Research Fellowship and the Sandler Translational Research Fellowship.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to R.C.F. ([rfoemke@phy.ucsf.edu](mailto:rfoemke@phy.ucsf.edu)).



## LETTERS

# BAI1 is an engulfment receptor for apoptotic cells upstream of the ELMO/Dock180/Rac module

Daeho Park<sup>1,2</sup>, Annie-Carole Tosello-Tramont<sup>2</sup>, Michael R. Elliott<sup>2</sup>, Mingjian Lu<sup>3,†</sup>, Lisa B. Haney<sup>2</sup>, Zhong Ma<sup>2</sup>, Alexander L. Klivanov<sup>4</sup>, James W. Mandell<sup>5</sup> & Kodi S. Ravichandran<sup>2,3</sup>

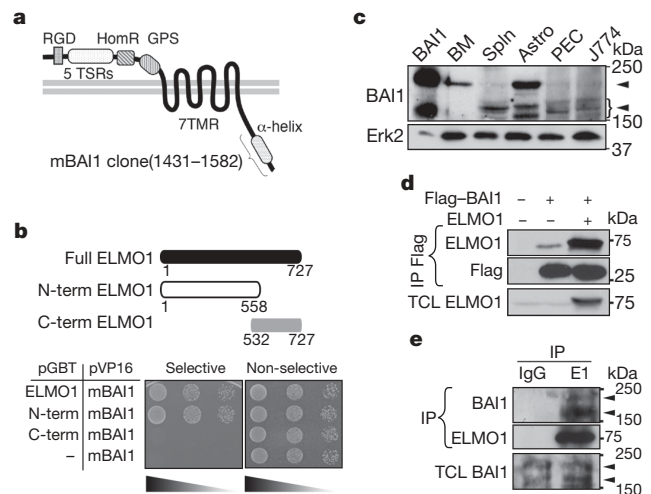
Engulfment and subsequent degradation of apoptotic cells is an essential step that occurs throughout life in all multicellular organisms<sup>1–3</sup>. ELMO/Dock180/Rac proteins are a conserved signalling module for promoting the internalization of apoptotic cell corpses<sup>4,5</sup>; ELMO and Dock180 function together as a guanine nucleotide exchange factor (GEF) for the small GTPase Rac, and thereby regulate the phagocyte actin cytoskeleton during engulfment<sup>4–6</sup>. However, the receptor(s) upstream of the ELMO/Dock180/Rac module are still unknown. Here we identify brain-specific angiogenesis inhibitor 1 (BAI1) as a receptor upstream of ELMO and as a receptor that can bind phosphatidylserine on apoptotic cells. BAI1 is a seven-transmembrane protein belonging to the adhesion-type G-protein-coupled receptor family, with an extended extracellular region<sup>7–9</sup> and no known ligands. We show that BAI1 functions as an engulfment receptor in both the recognition and subsequent internalization of apoptotic cells. Through multiple lines of investigation, we identify phosphatidylserine, a key 'eat-me' signal exposed on apoptotic cells<sup>10–13</sup>, as a ligand for BAI1. The thrombospondin type 1 repeats within the extracellular region of BAI1 mediate direct binding to phosphatidylserine. As with intracellular signalling, BAI1 forms a trimeric complex with ELMO and Dock180, and functional studies suggest that BAI1 cooperates with ELMO/Dock180/Rac to promote maximal engulfment of apoptotic cells. Last, decreased BAI1 expression or interference with BAI1 function inhibits the engulfment of apoptotic targets *ex vivo* and *in vivo*. Thus, BAI1 is a phosphatidylserine recognition receptor that can directly recruit a Rac–GEF complex to mediate the uptake of apoptotic cells.

Previous studies revealed two 'functional' regions within ELMO1 and its *Caenorhabditis elegans* homologue CED-12 during phagocytosis<sup>5,14–17</sup>. The amino-terminal 558 amino-acid residues (N-term) were necessary for targeting of the ELMO–Dock180 complex to the membrane<sup>14,17</sup>, whereas the carboxy-terminal 196 residues (C-term) were necessary for binding Dock180 and for optimal Rac activation<sup>15,16</sup>. Because the receptor(s) upstream of ELMO1 during engulfment were not known, we performed a yeast two-hybrid screen, with N-term as bait. After screening more than  $1.1 \times 10^7$  colonies from a mouse embryo library, followed by several subscreens for specificity, we identified a single membrane protein, BAI1.

BAI1 belongs to subgroup VII of the adhesion-type G-protein-coupled receptor (GPCR) family<sup>7–9</sup>, with extended extracellular termini containing multiple domains and motifs that are thought to function in cell–cell or cell–matrix interactions<sup>9</sup>. BAI1 (1,584 residues) has an 943-residue extracellular region, a seven-transmembrane 'heptahelical body' and a 392-residue cytoplasmic tail<sup>7,8</sup> (Fig. 1a). *bail* was initially cloned as a p53-regulated message in the brain<sup>7</sup> and received its name because an extracellular fragment inhibited

neovascularization in experimental angiogenesis<sup>7</sup>. However, no physiological ligands for BAI1 have been reported. The BAI1 fragment isolated in the two-hybrid screen (residues 1431–1582) was part of its cytoplasmic tail. Yeast transformants expressing BAI1<sup>1431–1582</sup> and either N-term or full-length ELMO1 (but not C-term) were able to grow under selective conditions (Fig. 1b). Binding of N-term to the cytoplasmic tail of BAI1 or to full-length BAI1 was also confirmed in mammalian cells (Supplementary Fig. 2, and data not shown).

Publicly available gene expression databases indicate the expression of *bail* outside the brain, and microarray analyses also reported *bail* expression in primary human monocytes and macrophages<sup>18</sup>. We detected *bail* mRNA and BAI1 protein at different levels in macrophage cell lines (J774 and RAW264.7) and primary tissues such as bone marrow and spleen (Fig. 1c and Supplementary Fig. 1). As reported previously<sup>7,8</sup>, endogenous BAI1 migrated at the predicted



**Figure 1 | Identification of BAI1 as an ELMO1-interacting protein.**

**a**, Diagram of BAI1. RGD, integrin-binding motif; HomR, hormone receptor; GPS, GPCR proteolytic site; 7TMR, seven-transmembrane receptor. **b**, A BAI1 clone was tested for interaction with the indicated ELMO1 constructs for growth on selective plates at tenfold dilutions. **c**, Immunoblotting for BAI1 expression shows a 160–170-kDa lower band (often a doublet) and a 220-kDa upper band<sup>7,8</sup>. BM, bone marrow; Spln, spleen; Astro, astrocytes; PEC, peritoneal exudate cells. **d**, Flag-BAI1 was transfected into LR73 cells, and simultaneous precipitation of endogenous ELMO1 (or transfected ELMO1; lane 3) was determined by anti-ELMO1 immunoblotting. IP, immunoprecipitation. TCL, total cell lysates. **e**, Mouse brain lysates were immunoprecipitated with ELMO1 antibody and simultaneous precipitation of endogenous BAI1 (arrowheads), assessed by immunoblotting. E1, ELMO1.

<sup>1</sup>Department of Cell Biology, <sup>2</sup>Carter Immunology Center, <sup>3</sup>Department of Microbiology, <sup>4</sup>Cardiovascular Division and <sup>5</sup>Department of Pathology, University of Virginia, Charlottesville, Virginia 22908, USA. <sup>†</sup>Present address: Institute for Diabetes, Obesity and Metabolism, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA.

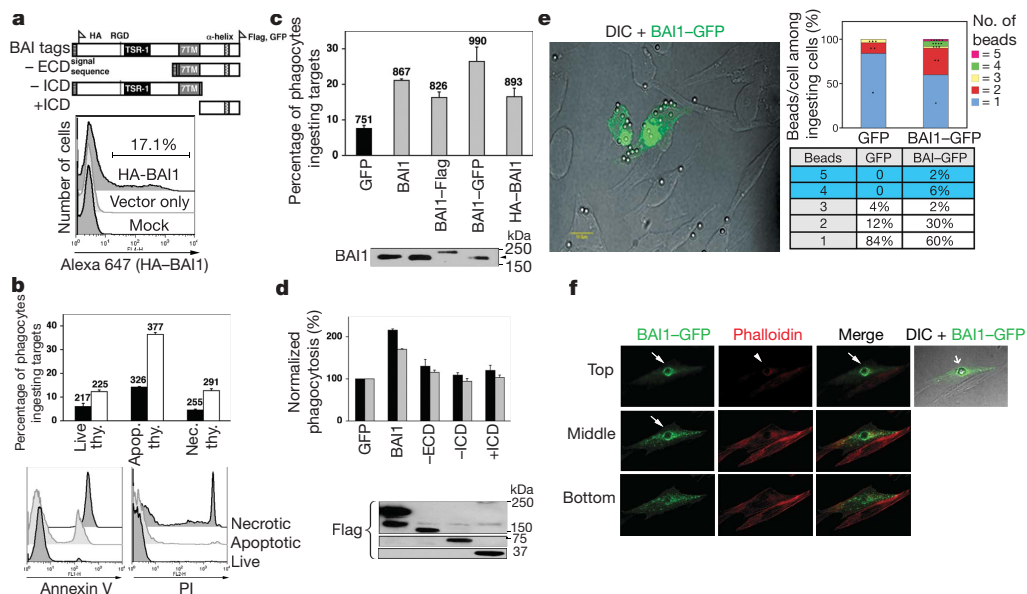
160–170 kDa (often as a doublet) and an upper 220-kDa band, with various intensities in different cell types (Fig. 1c). Endogenous ELMO1 precipitated together with overexpressed Flag-tagged BAI1, and association at native levels was also confirmed by co-precipitation of endogenous ELMO1 and BAI1 from mouse brain lysates (Fig. 1d, e).

We confirmed that an N-terminally haemagglutinin (HA)-tagged BAI1 (with the tag inserted after the leader sequence) was detectable on the cell surface by flow cytometry (Fig. 2a). Untagged BAI1 was also present in the membrane fractions (Supplementary Fig. 3). To address the role of BAI1 as an engulfment receptor, we generated J774 macrophage cells stably expressing the HA-tagged BAI1. In engulfment assays based on flow cytometry, expression of BAI1 enhanced the uptake of apoptotic thymocytes compared with controls (a fold increase of  $2.61 \pm 0.11$  (mean  $\pm$  s.d.),  $P < 0.0001$ ,  $n = 7$ ; Fig. 2b). The increased uptake by BAI1-expressing J774 cells occurred more towards apoptotic cells than towards necrotic and live cells (Fig. 2b). LR73 fibroblasts transiently transfected with BAI1 carrying different tags also enhanced the engulfment of apoptotic thymocytes (Supplementary Fig. 4) or fluorescently labelled 2- $\mu$ m carboxylate-modified beads, surrogate targets that mimic apoptotic cells<sup>19,20</sup> (a fold increase of  $2.62 \pm 0.20$  for green fluorescent protein (GFP)-labelled BAI1,  $P < 0.001$ ,  $n = 16$ ; Fig. 2c). Thus, the BAI1-dependent increased uptake occurred in both macrophages and non-professional phagocytes, with the effect being independent of the tag added. BAI1 lacking either the cytoplasmic or extracellular regions showed no such enhanced uptake, indicating a requirement for both regions (Fig. 2d). Expression of the BAI1 cytoplasmic region alone, capable of binding ELMO1, also had a minimal effect on engulfment over the control (Fig. 2d).

We then tested whether BAI1 promoted engulfment by enhanced binding of targets, by internalization, or by both methods. First, BAI1-GFP-transfected cells showed increased binding of targets (Fig. 2e); second, BAI1-expressing cells ingesting targets had a higher

mean fluorescence intensity (MFI) (where MFI reflects the target-derived fluorescence per cell<sup>4</sup>; Fig. 2b, c). When we quantified the efficiency of internalization (with Amnis ImageStream<sup>21</sup>, which combines flow cytometry with imaging of individual cells), BAI1-GFP-expressing cells exhibited enhanced engulfment, with two or more targets per phagocyte (40% versus 16% in the control; Fig. 2e). Moreover, only BAI1-GFP cells had four or five targets per phagocyte (Fig. 2e). BAI1-GFP was enriched at the phagocytic cup and its localization was correlated with polymerized actin at the same site, as determined by confocal microscopy (Fig. 2f). Thus, BAI1 contributes to both the binding and the engulfment of targets.

BAI1 contains five thrombospondin type 1 repeats (TSRs) in its extracellular region (Fig. 1a). Because TSRs can bind phosphatidylserine (PtdSer)<sup>22</sup>, we examined whether the TSRs of BAI1 participated in target recognition. Several lines of evidence collectively suggested a role for the TSRs of BAI1 in direct PtdSer recognition and in promoting the engulfment of apoptotic targets. First, the addition of annexin V, which binds PtdSer exposed on apoptotic cells, inhibited BAI1-mediated uptake in a dose-dependent manner (Fig. 3a). We then generated bacterial constructs with or without the TSRs (denoted RGD-TSR or RGD- $\Delta$ TSR, respectively; Fig. 3b). The addition of bacterially purified RGD-TSR protein inhibited BAI1-dependent uptake in a dose-dependent manner, whereas glutathione S-transferase (GST) alone or RGD- $\Delta$ TSR did not (Fig. 3c and Supplementary Fig. 5). Because the extracellular region of BAI1 also contains an RGD motif, which could potentially participate in integrin-dependent phagocyte–target interactions, we tested proteins with mutations in the RGD motif (RGE-TSR; Fig. 3b). RGE-TSR still inhibited uptake, but deletion of the TSRs (RGE- $\Delta$ TSR) abrogated the inhibition (Fig. 3c). The RGD motif in BAI1 was therefore dispensable for engulfment under these conditions, whereas the TSRs were essential.



**Figure 2 | BAI1 enhances uptake of apoptotic targets.** **a**, Top: diagram of BAI1 plasmid constructs used. Bottom: N-terminally HA-tagged BAI1 on the cell surface of transfected LR73 cells was assessed by flow cytometry with anti-HA antibody. ECD, extracellular domain; ICD, intracellular domain. **b**, J774 macrophage cells stably transfected with BAI1 (open columns) or control vector (filled columns) were incubated with live, necrotic or apoptotic thymocytes (thy.) and assessed for enhanced uptake. The numbers above the bars refer to MFI, indicative of target-derived fluorescence per phagocyte. **c**, LR73 cells were transiently transfected with tagged or untagged BAI1, and the uptake of 2- $\mu$ m carboxylate-modified beads or apoptotic thymocytes was assessed by a flow-cytometry-based engulfment assay (with MFI of cells ingesting targets shown). Expression of the transfected proteins was confirmed. **d**, BAI1 lacking the intracellular domain (ICD) or

extracellular domain (ECD) regions, or the ICD alone were tested in LR73 cells for BAI1-mediated increased uptake of apoptotic thymocytes (grey columns) and surrogate targets (black columns). **e**, Left: BAI1-GFP-expressing LR73 cells were incubated with unlabelled 2- $\mu$ m carboxylate beads; the enhanced binding of targets expressing BAI1-GFP is shown. DIC, differential interference contrast. Right: using Amnis Image Stream, the number of targets within the GFP-expressing or BAI1-GFP-expressing cells were counted and plotted. **f**, BAI1-GFP-expressing LR73 cells were incubated with 6- $\mu$ m carboxylate beads and stained with phalloidin; images from different confocal planes are shown to indicate the localization of BAI1-GFP to the phagocytic cup. BAI1 (arrow) and actin (arrowhead) are indicated. Error bars indicate s.d.

When we tested the direct binding of bacterially purified TSRs to apoptotic cells, RGD-TSR could decorate the surfaces of apoptotic cells, in a comparable manner to annexin V staining (Fig. 3d). In competition experiments, unlabelled soluble BAI1-TSR (but not those lacking the TSRs) inhibited the binding of fluorescent annexin V to surrogate targets (Fig. 3e) and apoptotic cells (Supplementary Fig. 6). Furthermore, LR73 cells expressing BAI1-GFP showed enhanced uptake of synthetic 2- $\mu$ m microbubbles containing phosphatidylserine (15% PtdSer and 85% phosphatidylcholine; Fig. 3g), with negligible uptake of phosphatidylcholine microbubbles.

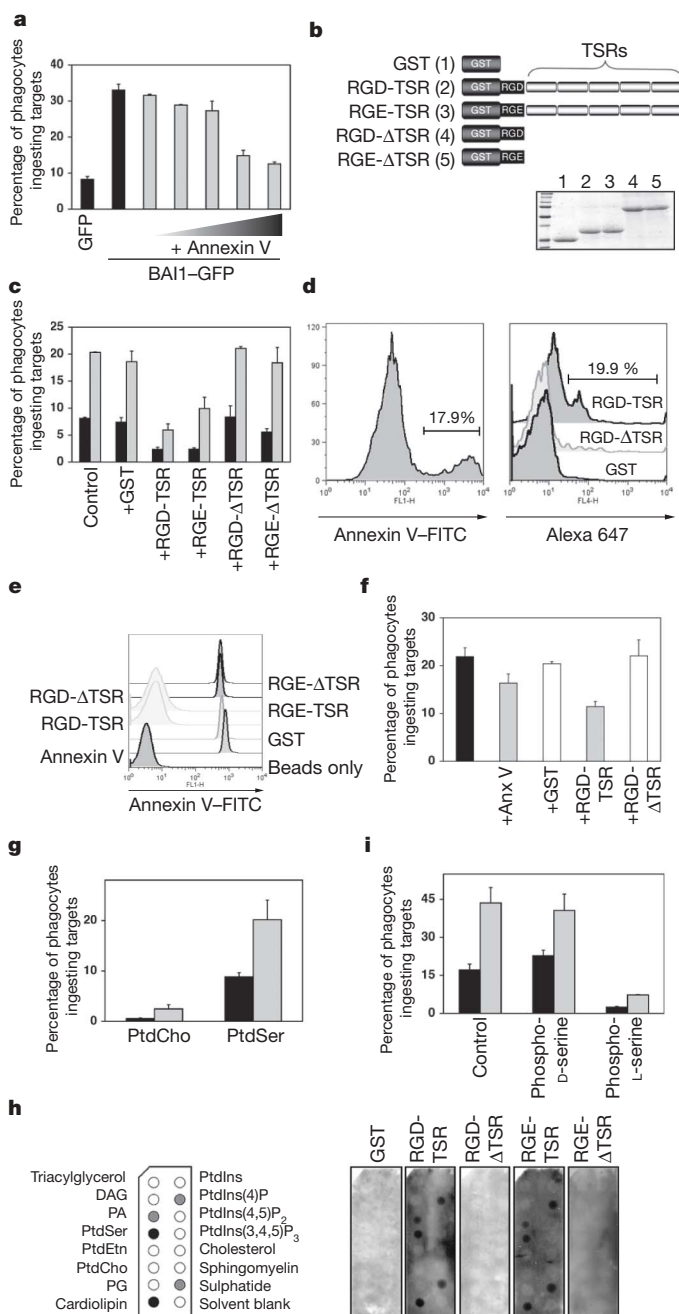
Purified proteins with BAI1-TSRs also bound to membrane strips spotted with various phospholipids (Fig. 3h). Under these conditions, the TSRs also showed binding to cardiolipin (a mitochondrial lipid), as well as weaker binding to phosphatidic acid and PtdIns(4)P, but not to other negatively charged lipids on the membrane (Fig. 3h). The binding to sulphatide served as a positive control<sup>23</sup>. This suggested a direct interaction between BAI1-TSRs and PtdSer in the absence of other cellular proteins. PtdSer recognition through

BAI1 seemed stereospecific, because the addition of phospho-L-serine, but not phospho-D-serine, strongly inhibited the BAI1-dependent enhanced uptake of apoptotic cells (Fig. 3i). Taken together, the above data showed BAI1 to be a new type of PtdSer-recognition receptor and implicated the TSRs of BAI1 in the direct interaction with phosphatidylserine.

We next addressed the requirement for ELMO1 binding to BAI1 in mediating signalling to promote engulfment. We first narrowed the ELMO1 binding to a short  $\alpha$ -helical stretch within the BAI1 cytoplasmic tail, which was necessary and sufficient (Supplementary Fig. 7). Mutation of three positive charges within this  $\alpha$ -helix (RKR $\rightarrow$ AAA) reproducibly reduced the binding of ELMO1 to BAI1 (Fig. 4a). Full-length BAI1 with this mutation failed to promote enhanced engulfment of apoptotic thymocytes (Fig. 4b). To test the role of ELMO1 in BAI1-mediated uptake directly, we performed stable knockdown of ELMO1 in J774 macrophages with short hairpin RNA (shRNA) targeted to *elmo1*. Transient expression of BAI1 promoted engulfment of apoptotic thymocytes in control shRNA cells but not in cells with ELMO1 knockdown (Fig. 4c). This suggested a requirement for ELMO1 expression for enhanced engulfment mediated by BAI1.

BAI1 formed a complex with Dock180, but this association was indirect because it required an interaction of BAI1 with ELMO1 (Fig. 4d). Guanine nucleotide exchange activity towards Rac was detectable in BAI1 precipitates when expressed with ELMO and Dock180 (data not shown). We then asked whether BAI1 expression would alter the levels of activated Rac (Rac-GTP) during phagocytosis. The basal level of Rac-GTP in BAI1-transfected LR73 cells was comparable to that in the mock-transfected cells. However, BAI1-expressing cells showed greater Rac-GTP after addition of targets (30 min), with the Rac-GTP levels returning to the basal state at later time points (Fig. 4e). These data suggested an inducible and transient Rac activation due to BAI1 during the recognition and engulfment of targets.

Whereas overexpression of BAI1 alone or of Dock180/ELMO proteins in LR73 cells increased engulfment, expression of all three proteins simultaneously caused the greatest increase in uptake of both apoptotic thymocytes and surrogate targets (Fig. 4f). This enhancement was lost when a mutant Dock180 (Dock-ISP)<sup>5</sup> that cannot activate Rac was expressed at the same time, and also resulted in inhibition of the uptake due to BAI1. Similarly, a mutant form of ELMO1 (T625) that can bind BAI1 but cannot link to Dock180 (ref. 5) also failed to promote uptake (Fig. 4f). Last, a dominant-negative



**Figure 3 | Recognition of phosphatidylserine by TSRs.** **a**, LR73 cells transfected with BAI1 were tested for engulfment of 2- $\mu$ m carboxylate beads in the presence or absence of annexin V (0.01–10 ng  $\mu$ l<sup>-1</sup>). **b**, Schematic diagram of constructs containing the TSRs and/or RGD motifs, and Coomassie staining of purified proteins. **c**, Addition of soluble RGD-TSR (10 ng  $\mu$ l<sup>-1</sup>), but not GST or  $\Delta$ TSR, blocked BAI1-mediated enhanced uptake in LR73 cells. Black columns, GFP; grey columns, BAI1. **d**, Apoptotic Jurkat cells were incubated with indicated soluble BAI1 fragments and their binding was assessed by flow cytometry. Binding of annexin V to the same apoptotic cells is shown for comparison. FITC, fluorescein isothiocyanate. **e**, Soluble RGD-TSR, RGE-TSR, RGD- $\Delta$ TSR and RGE- $\Delta$ TSR (2 ng  $\mu$ l<sup>-1</sup> each) were tested for blocking the binding of annexin V to carboxylate-modified beads by flow cytometry. **f**, Uptake of apoptotic thymocytes by J774 cells was tested in the presence of the indicated proteins (10 ng  $\mu$ l<sup>-1</sup> each). Anx, annexin. **g**, LR73 cells transfected with BAI1-GFP (grey columns) or control GFP alone (black columns) were tested for uptake of phosphatidylserine (PtdSer)-containing microbubbles or the control phosphatidylcholine (PtdCho) bubbles. **h**, Direct binding of soluble BAI1 proteins to phosphatidylserine and other phospholipids spotted on nitrocellulose membrane strips. DAG, diacylglycerol; PtdEtn, phosphatidylethanolamine. PA, phosphatidic acid, PG, phosphatidylglycerol. **i**, Stereospecificity of PtdSer recognition through BAI1-TSRs was tested by the addition of phospho-L-serine or phospho-D-serine during uptake of apoptotic thymocytes by control (black columns) or BAI1-overexpressing (grey columns) J774 cells. Error bars indicate s.d.



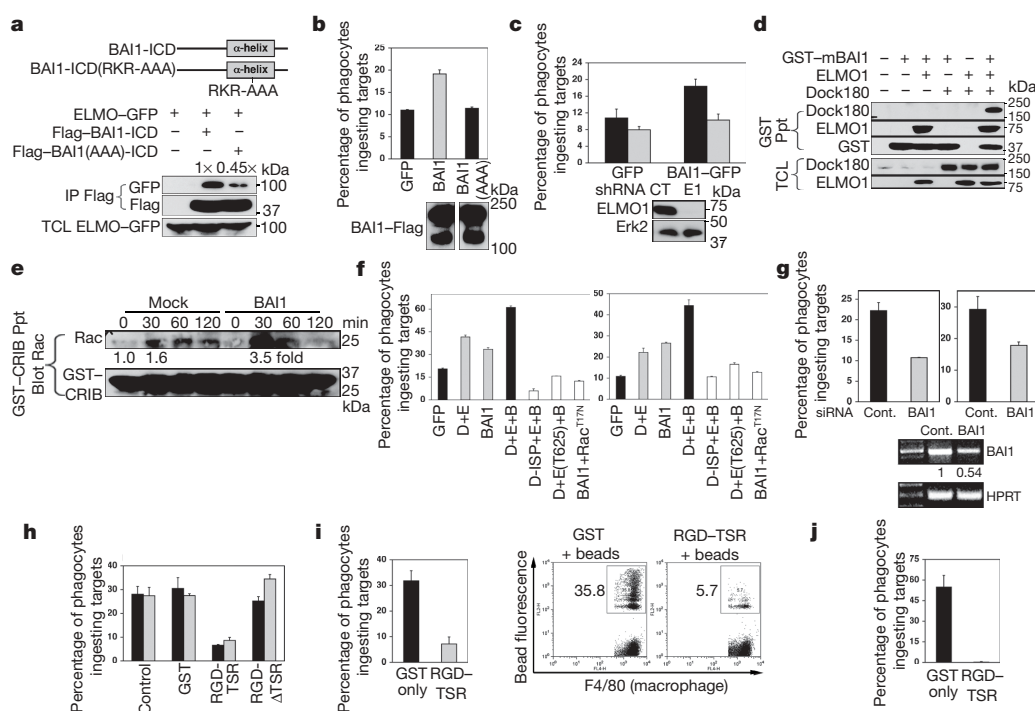
mutant of Rac ( $\text{Rac}^{\text{T17N}}$ ) also inhibited the uptake mediated by BAI1 (Fig. 4f). The augmented uptake due to simultaneous expression of BAI1, ELMO1 and Dock180 was due to the internalization of targets rather than to enhanced binding (not shown). Taken together, these data suggest that the intracellular tail of BAI1 mediates signalling through the ELMO/Dock180/Rac module to promote the engulfment of apoptotic cells.

Previous studies have suggested redundancy among engulfment receptors expressed on the surface of phagocytes. To test the relative importance for BAI1-mediated engulfment, we performed three types of analysis. First, short interfering RNA (siRNA)-mediated knockdown of BAI1 in primary mouse astrocytes resulted in a decreased uptake of apoptotic thymocytes or carboxylate beads (Fig. 4g). The efficiency of BAI1 knockdown (46%) was approximately correlated with the observed roughly 50% inhibition of engulfment. In J774 macrophages we could achieve only minimal knockdown of BAI1, but this was commensurate with a modest decrease in engulfment (Supplementary Fig. 8). The addition of soluble BAI1 TSRs with apoptotic thymocytes or carboxylate beads also blocked the engulfment of apoptotic cells by J774 cells (Fig. 3f) and primary mouse astrocytes (Fig. 4h). This suggested that BAI1-mediated recognition, under native conditions, does contribute to engulfment in macrophages.

To test the potential importance of BAI1-TSR-dependent recognition *in vivo*, we injected red-fluorescent-labelled carboxylate beads

or apoptotic thymocytes into the peritoneum of mice, with or without simultaneous injection of a soluble BAI1-TSR fragment or GST. After 20 min, we collected the peritoneal lavage cells and determined the fraction of F4/80-positive macrophages with engulfed targets. Simultaneous injection of BAI1-TSRs, but not that of GST, strongly inhibited the ability of peritoneal macrophages to engulf carboxylate-modified beads (Fig. 4i) or apoptotic thymocytes (Fig. 4j) and Supplementary Fig. 9). These data further support a role for BAI1 as an engulfment receptor in this *in vivo* model of engulfment.

Taken together, our data yield the following observations. Over the years, the receptor(s) that can directly engage PtdSer and how they may signal intracellularly have remained elusive. This work reveals BAI1 as one type of engulfment receptor that can stereospecifically engage PtdSer on apoptotic cells (through its TSRs). Our work also identifies a receptor upstream of the ELMO/Dock180/Rac signalling module, which has also remained unknown. This could be of great help in defining molecular events from PtdSer-dependent apoptotic cell recognition to responses within the phagocyte<sup>12</sup>. BAI1 is also a new type of engulfment receptor belonging to the adhesion-type GPCR family. Treatment of BAI1-transfected cells with pertussis toxin did not affect the enhanced uptake through BAI1, but BAI1 with gross deletion of the seven-transmembrane region failed to promote engulfment (data not shown). Thus, the specific role of the seven-transmembrane region of BAI1 remains to be determined.



**Figure 4 | BAI1 functions upstream of the ELMO/Dock180/Rac module.**

**a**, Top: diagram of the cytoplasmic tail of BAI1 and the RKR→AAA mutation. Binding of this BAI1 mutant to ELMO1 was assessed (relative densitometry signals for ELMO1 are indicated). **b**, Full-length BAI1 with the RKR→AAA mutation fails to enhance uptake of apoptotic thymocytes in LR73 cells. **c**, Top: control shRNA (black columns) or ELMO1 knockdown J774 cells (grey columns) were transfected with either BAI1-GFP or GFP alone, and the uptake of apoptotic thymocytes was assessed. Bottom: knockdown of ELMO1. **d**, 293T cells were transfected with the indicated combination of BAI1, ELMO1 and Dock180 plasmids, and complex formation was determined by immunoblotting. Ppt, precipitation. **e**, Top: Rac-GTP levels after incubation of targets with BAI1-expressing cells or controls. Bottom: densitometry for Rac-GTP signal and comparable GST-CRIB domain in the different lanes are shown. **f**, Engulfment of apoptotic thymocytes (left) or 2- $\mu\text{m}$  carboxylate beads (right) by LR73 cells transfected with BAI1-GFP (B) and the indicated combinations of ELMO1 (E), Dock180 (D), and the dominant-negative forms of ELMO1 (ET625),

Dock180 (D-ISP) or Rac ( $\text{Rac}^{\text{T17N}}$ ). Protein expression of the transfected plasmids was confirmed (data not shown). **g**, Top: knockdown of BAI1 in primary mouse astrocytes inhibits engulfment of apoptotic thymocytes (left) and surrogate targets (right). Bottom: the knockdown, as determined by *bai1* expression or control (cont.) hypoxanthine-guanine phosphoribosyltransferase (HPRT) message, and the efficiency of knockdown. **h**, Engulfment of apoptotic thymocytes (grey columns) or carboxylate beads (black columns) by primary astrocytes was assessed in the presence of the indicated soluble BAI1 proteins or GST alone ( $10 \text{ ng } \mu\text{l}^{-1}$  each). **i**, BAI1 TSR repeats or GST were injected together with red-labelled carboxylate beads into the peritoneum of C57/Bl6 mice (three mice per condition, two independent experiments). After 20 min, the fraction of F4/80-positive peritoneal macrophages ingesting targets was assessed (left) and a representative flow cytometry panel is also shown (right). **j**, As in **i**, except that LysM-Cre/YFP mice were used and the uptake of apoptotic thymocytes by YFP-positive cells (98% being F4/80-positive) was assessed. Error bars indicate s.d.

Last, the loss of BAI1 from gliomas seems to promote the formation of more aggressive glioblastomas<sup>7,8</sup>; conversely, overexpression of BAI1 inhibits the neo-vascularization of gliomas and inhibits tumour formation<sup>24</sup>. Because the natural function of BAI1 in tissues is not known, its identification as an engulfment receptor upstream of the ELMO/Dock180/Rac module could have implications for the treatment of glioblastomas.

## METHODS SUMMARY

**General.** Immortalized cell lines and primary cells were maintained as described previously<sup>5</sup>. All DNA constructs were generated by PCR-based methodology, with the fidelity and presence of appropriate mutations being confirmed by sequencing. The HF7C yeast strain (with His, Trp and Leu as selection markers)<sup>25</sup> was used for the yeast two-hybrid screen against a mouse embryo library. Immunoprecipitation and GST pull-down assays were used to detect protein–protein interactions<sup>20</sup>. Protein expressions and the presence of bound proteins were confirmed by western blotting. Transiently transfected siRNA and stably expressed shRNA were used to knock down the level of BAI1 or ELMO1 proteins. RT–PCR was used to detect the level of mRNA of a specific gene.

**In vitro phagocytosis assay.** Various cell lines (LR73 Chinese hamster ovary cells, J774 macrophages and NIH 3T3 cells) and primary cells (mouse astrocytes) were transiently or stably transfected before the phagocytosis assays<sup>20</sup>. Phagocytic cells were incubated with fluorescently labelled targets such as apoptotic thymocytes, 2- $\mu$ m carboxylate-modified latex beads, or 2- $\mu$ m phospholipid containing microbubbles. After 2 h, the cells were extensively washed with cold PBS, treated with trypsin and resuspended in cold medium containing 1% NaN<sub>3</sub>, and analysed by flow cytometry as described previously<sup>20</sup>.

**In vivo engulfment assay.** Surrogate beads or apoptotic thymocytes stained with 5- (and 6)-carboxytetramethylrhodamine (TAMRA) were mixed with soluble GST or RGD-TSR protein. The mixture was then injected into the peritoneum of either wild-type C57/Bl6 mice or a LysM-Cre/yellow fluorescent protein (YFP) reporter mouse line (in which 98% of YFP-positive cells in the peritoneum were also F4/80 positive). Peritoneal exudate cells were collected and either F4/80-positive or YFP-positive cells were gated and fluorescence-activated cell sorting analyses were performed to determine the effects of soluble proteins on phagocytosis *in vivo*.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 26 July; accepted 1 October 2007.

Published online 24 October 2007.

1. Savill, J., Dransfield, I., Gregory, C. & Haslett, C. A blast from the past: clearance of apoptotic cells regulates immune responses. *Nature Rev. Immunol.* **2**, 965–975 (2002).
2. Henson, P. M. Engulfment: ingestion and migration with Rac, Rho and TRIO. *Curr. Biol.* **15**, R29–R30 (2005).
3. Gregory, C. D. & Brown, S. B. Apoptosis: eating sensibly. *Nature Cell Biol.* **7**, 1161–1163 (2005).
4. Gumienny, T. L. *et al.* CED-12/ELMO, a novel member of the crkl/Dock180/Rac pathway, is required for phagocytosis and cell migration. *Cell* **107**, 27–41 (2001).
5. Brugnera, E. *et al.* Unconventional Rac-GEF activity is mediated through the Dock180-ELMO complex. *Nature Cell Biol.* **4**, 574–582 (2002).
6. Lu, M. & Ravichandran, K. S. Dock180-ELMO cooperation in Rac activation. *Methods Enzymol.* **406**, 388–402 (2006).
7. Nishimori, H. *et al.* A novel brain-specific p53-target gene, BAI1, containing thrombospondin type 1 repeats inhibits experimental angiogenesis. *Oncogene* **15**, 2145–2150 (1997).
8. Kaur, B., Brat, D. J., Devi, N. S. & Van Meir, E. G. Vasculostatin, a proteolytic fragment of brain angiogenesis inhibitor 1, is an antiangiogenic and antitumorigenic factor. *Oncogene* **24**, 3632–3642 (2005).
9. Bjarnadottir, T. K. *et al.* The human and mouse repertoire of the adhesion family of G-protein-coupled receptors. *Genomics* **84**, 23–33 (2004).

10. Fadok, V. A. *et al.* Exposure of phosphatidylserine on the surface of apoptotic lymphocytes triggers specific recognition and removal by macrophages. *J. Immunol.* **148**, 2207–2216 (1992).
11. Henson, P. M. & Hume, D. A. Apoptotic cell removal in development and tissue homeostasis. *Trends Immunol.* **27**, 244–250 (2006).
12. Kiss, R. S., Elliott, M. R., Ma, Z., Marcel, Y. L. & Ravichandran, K. S. Apoptotic cells induce a phosphatidylserine-dependent homeostatic response from phagocytes. *Curr. Biol.* **16**, 2252–2258 (2006).
13. Wu, Y., Tibrewal, N. & Birge, R. B. Phosphatidylserine recognition by phagocytes: a view to a kill. *Trends Cell Biol.* **16**, 189–197 (2006).
14. Grimsley, C. M. *et al.* Dock180 and ELMO1 proteins cooperate to promote evolutionarily conserved Rac-dependent cell migration. *J. Biol. Chem.* **279**, 6087–6097 (2004).
15. Lu, M. *et al.* PH domain of ELMO functions in trans to regulate Rac activation via Dock180. *Nat. Struct. Mol. Biol.* **11**, 756–762 (2004).
16. Lu, M. *et al.* A steric-inhibition model for regulation of nucleotide exchange via the Dock180 family of GEFs. *Curr. Biol.* **15**, 371–377 (2005).
17. deBakker, C. D. *et al.* Phagocytosis of apoptotic cells is regulated by a UNC-73/TRIO-MIG-2/RhoG signaling module and armadillo repeats of CED-12/ELMO. *Curr. Biol.* **14**, 2208–2216 (2004).
18. Cho, H. *et al.* Induction of dendritic cell-like phenotype in macrophages during foam cell formation. *Physiol. Genomics* **29**, 149–160 (2007).
19. Erwig, L. P. *et al.* Differential regulation of phagosome maturation in macrophages and dendritic cells mediated by Rho GTPases and ezrin-radixin-moesin (ERM) proteins. *Proc. Natl Acad. Sci. USA* **103**, 12825–12830 (2006).
20. Tosello-Tramont, A. C. *et al.* Identification of two signaling submodules within the Crkl/ELMO/Dock180 pathway regulating engulfment of apoptotic cells. *Cell Death Differ.* **14**, 963–972 (2007).
21. Adang, L. A., Parsons, C. H. & Kedes, D. H. Asynchronous progression through the lytic cascade and variations in intracellular viral loads revealed by high-throughput single-cell analysis of Kaposi's sarcoma-associated herpesvirus infection. *J. Virol.* **80**, 10073–10082 (2006).
22. Manodori, A. B., Barabino, G. A., Lubin, B. H. & Kuypers, F. A. Adherence of phosphatidylserine-exposing erythrocytes to endothelial matrix thrombospondin. *Blood* **95**, 1293–1300 (2000).
23. Guo, N. H. *et al.* Heparin- and sulfatide-binding peptides from the type I repeats of human thrombospondin promote melanoma cell adhesion. *Proc. Natl Acad. Sci. USA* **89**, 3040–3044 (1992).
24. Kang, X. *et al.* Antiangiogenic activity of BAI1 *in vivo*: implications for gene therapy of human glioblastomas. *Cancer Gene Ther.* **13**, 385–392 (2006).
25. Lindsay, M. E., Holaska, J. M., Welch, K., Paschal, B. M. & Macara, I. G. Ran-binding protein 3 is a cofactor for Crm1-mediated nuclear protein export. *J. Cell Biol.* **153**, 1391–1402 (2001).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. Casanova, C. Grimsley and members of the Ravichandran laboratory for suggestions and for critical reading of the manuscript. This work was supported by grants from the National Institute of General Medical Sciences (to K.S.R.). M.R.E. was supported by a postdoctoral fellowship from the American Cancer Society, and Z.M. was supported by a postdoctoral fellowship through the BioDefense Training grant (NIH).

**Author Contributions** D.P. performed and analysed most of the experiments in this study. A.C.T. performed Amnis ImageStream studies and the confocal microscopy analyses. M.R.E. helped with RT–PCR analyses and the *in vivo* mouse studies with BAI1-TSR. M.L. generated ELMO1 knockdown 774 cells. L.B.H. analysed the GEF activity associated with BAI1 and provided technical help in other parts of the manuscript. Z.M. performed the lipid membrane strip binding assay of the TSR protein. A.L.K. generated 2- $\mu$ m lipid vesicles. J.W.S. provided primary astrocytes, technical expertise and critical intellectual input with these studies. K.S.R. provided overall coordination with respect to conception, design and supervision of the study. K.S.R. also wrote the manuscript with comments from co-authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to K.S.R. ([ravi@virginia.edu](mailto:ravi@virginia.edu)).

## METHODS

**Cell culture and transfections.** 293T cells and primary astrocytes were maintained in DMEM containing 10% FBS and 1% penicillin–streptomycin–glutamine<sup>5</sup>. LR73 Chinese hamster ovary cells, J774 macrophages and NIH 3T3 fibroblasts were cultured as described previously<sup>20</sup>. Plasmid transfections into 293T cells were performed with calcium phosphate (Promega), and LR73 and NIH/3T3 cell transfections were performed with Lipofectamine 2000 (Invitrogen) in accordance with the manufacturer's instructions. Nucleofection was used to introduce plasmids and siRNA into J774 cells (Amata kit V, programme T-20) and primary astrocytes (Amata astrocyte kit, programme T-20).

**Yeast two-hybrid screen.** HF7C yeast strain (His<sup>−</sup>, Trp<sup>−</sup>, Leu<sup>−</sup>) was used to screen a mouse seven-day embryonic library<sup>25</sup> using pGBT10-N-term ELMO1 (residues 1–558) as bait. After screening of  $1.1 \times 10^7$  colonies for growth on selective SCM plates (Trp<sup>−</sup>, Leu<sup>−</sup>, His<sup>−</sup> with 5 mM 3-amino-1,2,4-triazole), and additional specificity/selection steps, the library plasmids were rescued and sequenced.

**Plasmids and mutagenesis.** ELMO, Dock180 and Rac constructs used in this study have been described previously<sup>5,14</sup>. mBAI1,  $\alpha$ -helix and non- $\alpha$ -helix constructs were generated by a PCR-based strategy from the yeast clone identified in the two-hybrid screen. Full-length pcDNA3.1(+)-BAI1 was a gift from T. Tokino. The various BAI1 constructs were generated by a PCR-based strategy. Flag and GFP sequence tags in the pEBB-BAI1–Flag and pEBB-BAI1–GFP constructs were inserted at the C-terminal end of BAI1; pEBB-HA–BAI1 construct was generated by inserting a tandem HA sequence tag near the N terminus, after the leader sequence (at the 50th amino acid residue of BAI1). The pEBB-BAI1(without ECD)–Flag construct contained residues 1–50 fused to residues 951–1584, the pEBB-BAI1(without ICD)–Flag construct contained residues 1–1201, and the Flag–BAI1(ICD) construct contained residues 1192–1584. For bacterial production of GST–RGD–TSR, the sequence encoding residues 202–585 of BAI1 were introduced into the pGEX-4T2 vector. GST–RGD– $\Delta$ TSR contained residues 202–263 fused to residues 577–585 of BAI1. The RGD sequence of BAI1 (residues 231–233 of BAI1) was mutated to generate GST–RGE–TSR and GST–RGE– $\Delta$ TSR. All constructs were sequenced to confirm fidelity and the presence of the appropriate mutations.

**Immunoprecipitation and immunoblotting.** At 24–48 h after transfection of plasmids into 293T, LR73 or J774 cells, the cells were lysed and precipitated with the indicated antibodies or glutathione–Sepharose beads. Brain tissue was lysed and immunoprecipitated with anti-ELMO1 antibody conjugated with Protein A/G-agarose (Santa Cruz Biotechnology). Bound proteins were assessed by immunoblotting for the respective tags or endogenous proteins. The anti-ELMO1 antibody has been described previously<sup>5</sup>. The BAI1 antibody was obtained from Orbigen.

**Immunofluorescence staining.** NIH 3T3 or LR73 cells were plated on Labtek II slides and transfected with Lipofectamine 2000. At harvest, slides were washed with PBS, fixed in 3% formaldehyde and permeabilized with 0.2% Triton X-100. Cells were then blocked with clarified milk and subsequently stained with appropriate antibodies. Immunostaining was performed sequentially with IgG-specific secondary antibodies as described previously<sup>20</sup>.

**Quantitative RT–PCR.** Total RNA from cell lines, tissues or primary cells was extracted with the use of an RNeasy Mini Kit (Qiagen) in accordance with the manufacturer's instructions. Complementary DNA was generated from 1.0  $\mu$ g of total RNA with Superscript III (Invitrogen) in accordance with the protocol. In all, 35 PCR cycles were performed for the amplification of mouse BAI1 with 1  $\mu$ l of the 20  $\mu$ l reaction mixture of cDNA.

**Phagocytosis assay.** LR73 or J774 cells were transiently transfected in triplicate with the indicated plasmids in a 24-well plate. At 24 h after transfection, engulfment assays were performed essentially as described previously<sup>20</sup>, with either TAMRA-labelled apoptotic thymocytes or 2- $\mu$ m carboxylate-modified red fluorescent beads (Invitrogen) that mimic the negative charge on apoptotic cells and can serve as a simplified target<sup>12,19</sup>. After being washed ten times with cold PBS, the cells were treated with trypsin, resuspended in cold medium (with 1% NaN<sub>3</sub>) and analysed by two-colour flow cytometry. The transfected cells were recognized by their GFP fluorescence; targets (carboxylate-modified red fluorescent beads, apoptotic thymocytes or lipid vesicles) were recognized by red fluorescence. Forward and side-scatter parameters were used to distinguish unbound targets from cells. The data were analysed with FlowJo software.

Most double-positive cells scored in this flow-cytometry-based assay were targets engulfed by transfected cells or targets being engulfed, and not targets loosely bound to the cell surface (because these were removed during the extensive washing and trypsin treatment before flow cytometry).

The MFI in the red channel of the cells taking up targets gave an estimate of the efficiency of uptake (proportional to the number of particles taken up). To quantify the number of targets engulfed per cell, the images of the phagocytes as they were analysed by flow cytometry were obtained with Amnis ImageStream. Cell-surface staining of phagocytes for class I MHC expression was used to focus on the internalized targets and counting of targets was performed on a per-cell basis.

Red fluorescent microbubbles (2  $\mu$ m) containing phosphatidylcholine (100%) and phosphatidylserine (molar ratio of phosphatidylcholine to phosphatidylserine 85:15) were washed five times with PBS and resuspended with  $\alpha$ -MEM supplemented with 2% FBS and 0.2% penicillin–streptomycin–glutamine. Vesicles ( $4 \times 10^6$ ) were added to each well of a 24-well plate; each well was then filled with the same medium. The 24-well plate was sealed, flipped over, and incubated at 37 °C for 2 h; the uptake was analysed by flow cytometry.

For *in vivo* analysis of BAI1 function, 10<sup>7</sup> TAMRA-stained apoptotic thymocytes or 2  $\mu$ l of 2- $\mu$ m carboxylate beads (red labelled) were mixed with 3  $\mu$ g of GST or RGD–TSR proteins in 300  $\mu$ l of PBS. The 300  $\mu$ l mixtures of targets were injected into the peritoneum of either a wild-type or a LysM-Cre/YFP reporter mouse. At 20 min after injection, peritoneal exudate cells were collected. The cells were resuspended in 400  $\mu$ l of PBS containing 1% NaN<sub>3</sub>. For wild-type C57Bl/6 mice that were used for the bead injections, to block Fc receptors, CD16/32 antibody was treated for 15 min and then the cells were stained with fluorescently labelled F4/80 antibody for 15 min. For LysM-Cre/YFP mice, the more than 98% YFP-positive cells in the peritoneum were F4/80-positive and CD11b-positive (and were CD3-negative and B220-negative). Either YFP-positive or F4/80-positive cells were gated to evaluate the fraction of macrophages ingesting red-labelled targets.

**Binding assay of soluble TSRs.** Jurkat cells ( $2.5 \times 10^6$ ) were washed with PBS and resuspended with 10 ml of X-Vivo10 medium in the 100-mm Petri dishes. The Jurkat cells were irradiated with ultraviolet and incubated at 37 °C for 2.5 h. The cells were washed twice with cold PBS and then resuspended in annexin V binding buffer containing 10 mM HEPES–NaOH pH 7.4, 140 mM NaCl, 2.5 mM CaCl<sub>2</sub> at a concentration of  $10^5$  cells ml<sup>−1</sup>. Bacterially produced soluble TSR proteins (1  $\mu$ g) were incubated with 500  $\mu$ l of apoptotic Jurkat cell suspension for 15 min at room temperature (20 °C). After incubation, the cells were washed twice and resuspended with annexin V binding buffer. The presence of TSR proteins on apoptotic Jurkat cells was detected with anti-GST antibody (clone Z5; Santa Cruz) followed by Alexa 647-conjugated anti-IgG secondary antibody (Molecular Probes) and flow cytometry. The competitive binding of soluble TSRs with annexin V (Becton Dickinson) to 6- $\mu$ m carboxylate beads (Polysciences) or apoptotic Jurkat cells was performed as described above except that 1  $\mu$ g of the indicated proteins was incubated with the apoptotic Jurkat cells first for 15 min, followed by the addition of 5  $\mu$ l of FITC-conjugated annexin V for a further 15 min, and subsequent flow cytometry.

For inhibition assays, annexin V or the BAI1 recombinant fragments (10 ng  $\mu$ l<sup>−1</sup>) were added to 10<sup>7</sup> apoptotic thymocytes and incubated for 15 min at room temperature. The mixture of proteins and apoptotic thymocytes was then added to J774 cells on a 24-well plate and analysed in the engulfment assay.

**Binding of TSRs to lipids.** The membrane lipid strips (Echelon Biosciences, Inc.) were blocked in 3% fatty-acid-free BSA in TBS-T containing 50 mM Tris–HCl pH 8.0, 150 mM NaCl and 0.1% Tween-20 for 1 h at room temperature in the dark. The membrane was then incubated at 4 °C with gentle agitation in the same solution containing the indicated purified proteins (each at 0.5  $\mu$ g ml<sup>−1</sup>). The membranes were washed four times in TBS-T and then incubated for 1 h with anti-GST antibody (clone Z5; Santa Cruz) followed by routine immunoblotting steps.

**CRIB pull-down assay.** LR73 cells were transfected with BAI1 using Lipofectamine 2000. At 24 h after transfection, the cells were incubated with 2- $\mu$ m carboxylate beads for 30, 60 and 120 min. The cells were then washed with cold PBS twice and lysed. The cell lysates were incubated with 30  $\mu$ g of GST–CRIB for 1 h at 4 °C. The precipitation of Rac by GST–CRIB was detected by immunoblotting.



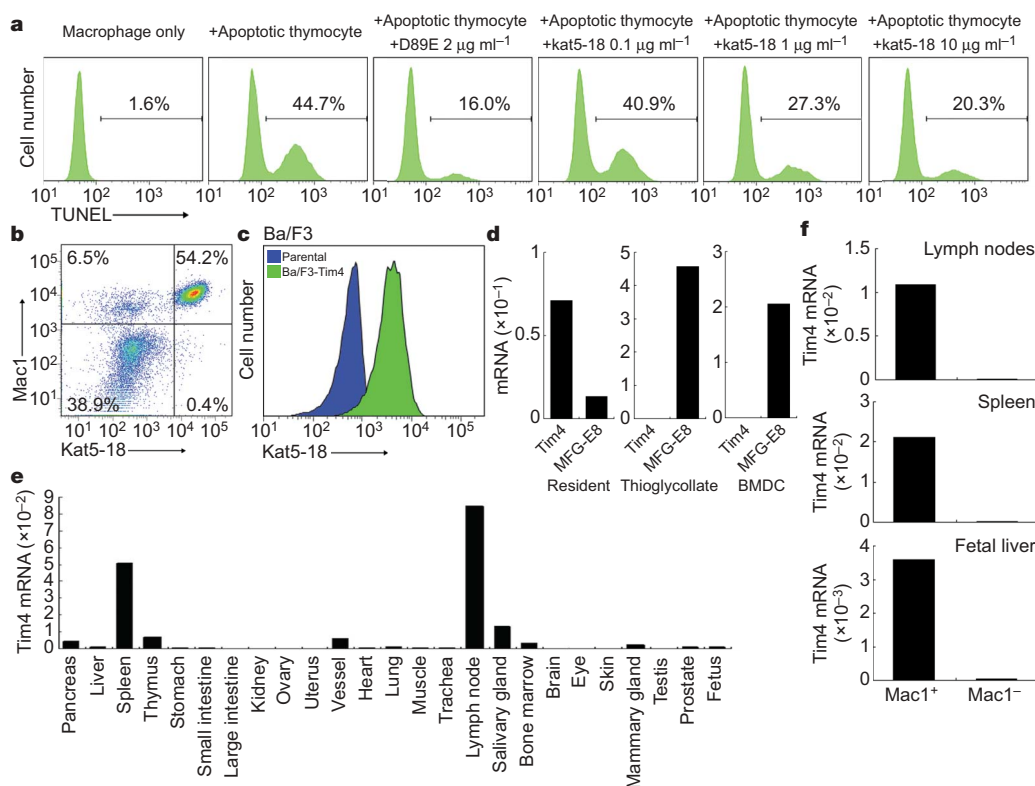
# Identification of Tim4 as a phosphatidylserine receptor

Masanori Miyanishi<sup>1,2</sup>, Kazutoshi Tada<sup>2†</sup>, Masato Koike<sup>3</sup>, Yasuo Uchiyama<sup>3</sup>, Toshio Kitamura<sup>4</sup> & Shigekazu Nagata<sup>1,2,5</sup>

In programmed cell death, a large number of cells undergo apoptosis, and are engulfed by macrophages to avoid the release of noxious materials from the dying cells<sup>1,2</sup>. In definitive erythropoiesis, nuclei are expelled from erythroid precursor cells and are engulfed by macrophages. Phosphatidylserine is exposed on the surface of apoptotic cells<sup>3</sup> and on the nuclei expelled from erythroid precursor cells<sup>4</sup>; it works as an 'eat me' signal for phagocytes<sup>5,6</sup>. Phosphatidylserine is also expressed on the surface of exosomes involved in intercellular signalling<sup>7</sup>. Here we established a library of hamster monoclonal antibodies against mouse peritoneal macrophages, and found an antibody that strongly inhibited the phosphatidylserine-dependent engulfment of apoptotic cells. The antigen recognized by the antibody was identified by expression cloning as a type I transmembrane protein called Tim4 (T-cell immunoglobulin- and mucin-domain-containing molecule; also known as Timd4)<sup>8</sup>. Tim4 was expressed in Mac1<sup>+</sup> cells in various mouse tissues, including spleen, lymph nodes and fetal

liver. Tim4 bound apoptotic cells by recognizing phosphatidylserine via its immunoglobulin domain. The expression of Tim4 in fibroblasts enhanced their ability to engulf apoptotic cells. When the anti-Tim4 monoclonal antibody was administered into mice, the engulfment of apoptotic cells by thymic macrophages was significantly blocked, and the mice developed autoantibodies. Among the other Tim family members, Tim1, but neither Tim2 nor Tim3, specifically bound phosphatidylserine. Tim1- or Tim4-expressing Ba/F3 B cells were bound by exosomes via phosphatidylserine, and exosomes stimulated the interaction between Tim1 and Tim4. These results indicate that Tim4 and Tim1 are phosphatidylserine receptors for the engulfment of apoptotic cells, and may also be involved in intercellular signalling in which exosomes are involved.

Caspase-activated DNase (CAD)-deficient cells do not undergo apoptotic DNA fragmentation, but their DNA is degraded in phagocytes after they are engulfed<sup>9</sup>. We used this knowledge to assay the



**Figure 1 | Tim4 in peritoneal macrophages.** **a**, Mouse peritoneal cells were used for phagocytosis with D89E<sup>10</sup> or Kat5-18, stained for TUNEL, analysed by flow cytometry, and the percentage of TUNEL<sup>+</sup> macrophages determined. Average values of two experiments are shown. **b**, Peritoneal macrophages incubated with biotin-Kat5-18, followed with FITC-streptavidin and APC-anti-Mac1. **c**, Parental and Tim4-expressing Ba/F3 analysed by flow cytometry with Kat5-18. **d**, mRNA levels of Tim4 and MFG-E8 in resident or thioglycollate-elicited peritoneal macrophages, and bone marrow-derived immature dendritic cells, determined by real-time PCR. **e**, Tim4 mRNA levels in mouse tissues, determined by real-time PCR. **f**, Real-time PCR for Tim4 mRNA in Mac1<sup>+</sup> and Mac1<sup>-</sup> cells from lymph nodes, spleen and fetal liver.

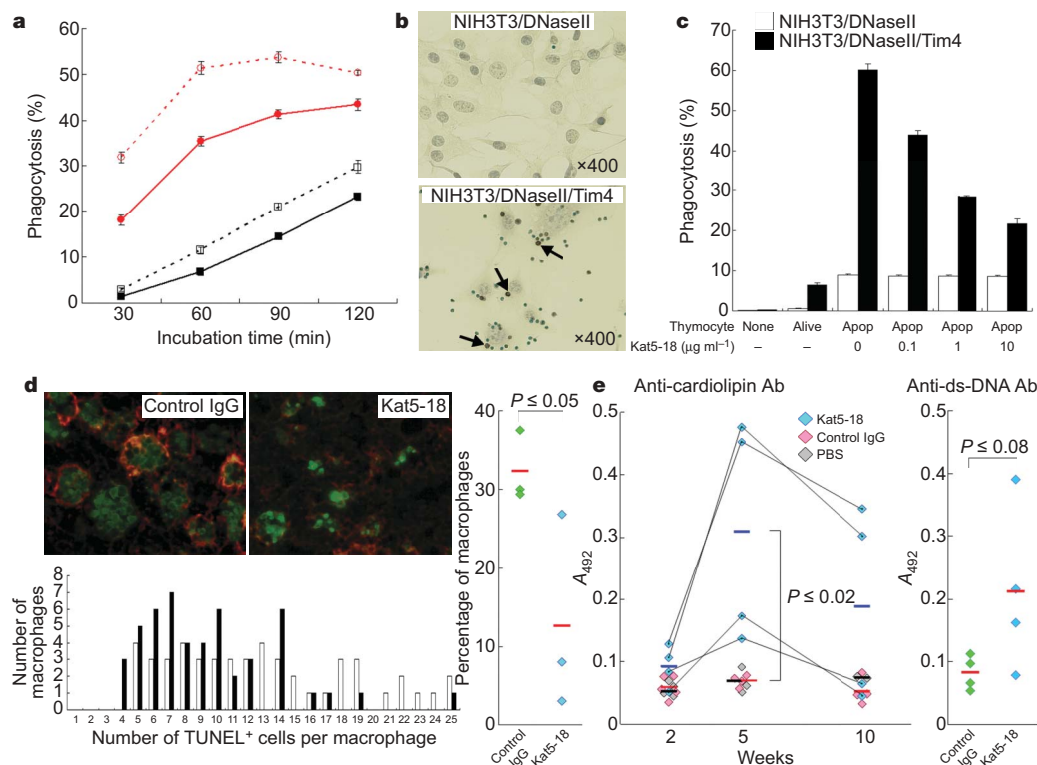
<sup>1</sup>Department of Medical Chemistry, Graduate School of Medicine, Kyoto University, Yoshida, Sakyo-ku, Kyoto 606-8501, Japan. <sup>2</sup>Department of Genetics, <sup>3</sup>Department of Cell Biology and Neuroscience, Osaka University Medical School, Osaka 565-0871, Japan. <sup>4</sup>Division of Cellular Therapy, The Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108-8639, Japan. <sup>5</sup>Solution Oriented Research for Science and Technology, Japan Science and Technology Corporation, Kyoto 606-8501, Japan. <sup>†</sup>Present address: Mitsubishi UFJ Research and Consulting, Corporate Strategy Consulting Department I, 2-5-8 Emabashi, Chou-ku, Osaka 541-8512, Japan

engulfment of apoptotic cells, and identified MFG-E8 (milk fat globule EGF factor 8) that stimulates their engulfment<sup>10</sup>. MFG-E8 is expressed in thioglycollate-elicited peritoneal macrophages, tingible-body macrophages of lymph nodes and spleen, Langerhans cells and mammary epithelial cells<sup>10–13</sup>. Hu *et al.*<sup>14</sup> reported that naïve peritoneal macrophages engulf apoptotic cells in a phosphatidylserine (PS)-dependent manner. We confirmed this with our assay system, which uses *CAD*<sup>-/-</sup> thymocytes as prey (Fig. 1a). Peritoneal macrophages efficiently engulfed apoptotic cells, and this was inhibited by the D89E mutant of MFG-E8 that masks PS<sup>10</sup>. However, peritoneal macrophages expressed little MFG-E8 (see below), and this MFG-E8 deficiency did not affect the ability of these macrophages to engulf apoptotic cells (data not shown). To identify the molecules involved in this engulfment, Armenian hamsters were immunized with mouse peritoneal cells, and hybridomas were prepared. Among 1,200 hybridomas, one monoclonal antibody (Kat5-18) dose-dependently inhibited the phagocytosis of apoptotic cells by peritoneal macrophages (Fig. 1a). Most *Mac1*<sup>+</sup> peritoneal cells were stained by Kat5-18 (Fig. 1b), indicating that peritoneal macrophages expressed its antigen.

To identify the antigen for Kat5-18, we used retrovirus-mediated expression cloning<sup>15</sup>. A complementary DNA library was constructed with messenger RNA from mouse peritoneal cells, converted to retroviruses, and used to infect mouse Ba/F3 cells. After 2 days in culture, transformants labelled by Kat5-18 (less than 0.1%) were sorted using a cell sorter, expanded in culture, and sorted again. This cycle of sorting and expansion was repeated three times until most cells

stained with Kat5-18. The sorted Kat5-18<sup>+</sup> cells carried 2–3 different integrated cDNAs, and *Tim4* cDNA was found in two independent experiments. When *Tim4* cDNA was introduced into Ba/F3, the transformants stained strongly with Kat5-18 (Fig. 1c), confirming that *Tim4* was the antigen recognized by Kat5-18. Real-time PCR indicated that *Tim4* mRNA was expressed in resident, but not thioglycollate-elicited, peritoneal cells (Fig. 1d), in sharp contrast with MFG-E8, which was expressed in thioglycollate-elicited but weakly in resident peritoneal cells. A similar situation was observed with bone marrow-derived immature dendritic cells, which expressed a high level of MFG-E8, but not *Tim4*. Other major tissues that expressed *Tim4* included the spleen, thymus, lymph nodes and salivary glands (Fig. 1e). When cells from the spleen, lymph nodes and fetal liver were sorted into *Mac1*<sup>+</sup> and *Mac1*<sup>-</sup> cells using MACS (magnetic cell sorting), *Tim4* mRNA was found only in *Mac1*<sup>+</sup> cells (Fig. 1f). Flow cytometry indicated that about 20% of *Mac1*<sup>+</sup> cells in the lymph nodes expressed *Tim4* (Supplementary Fig. 1).

DNase II degrades DNA of apoptotic cells in lysosomes after phagocytes engulf them<sup>16</sup>. To examine the role of *Tim4* in the engulfment of apoptotic cells, a fibroblast cell line expressing *Tim4* was established with NIH3T3 expressing DNase II<sup>10</sup> (NIH3T3/DNase II/*Tim4*) (Supplementary Fig. 2). The ability of the parental NIH3T3/DNase II to engulf apoptotic cells was weak: when incubated with apoptotic *CAD*<sup>-/-</sup> thymocytes for 120 min, only 20% of the cells engulfed them (Fig. 2a). On the other hand, more than 50% of NIH3T3/DNase II/*Tim4* engulfed apoptotic cells within 60 min. This effect of *Tim4* was confirmed by microscopic observation. That is, when apoptotic



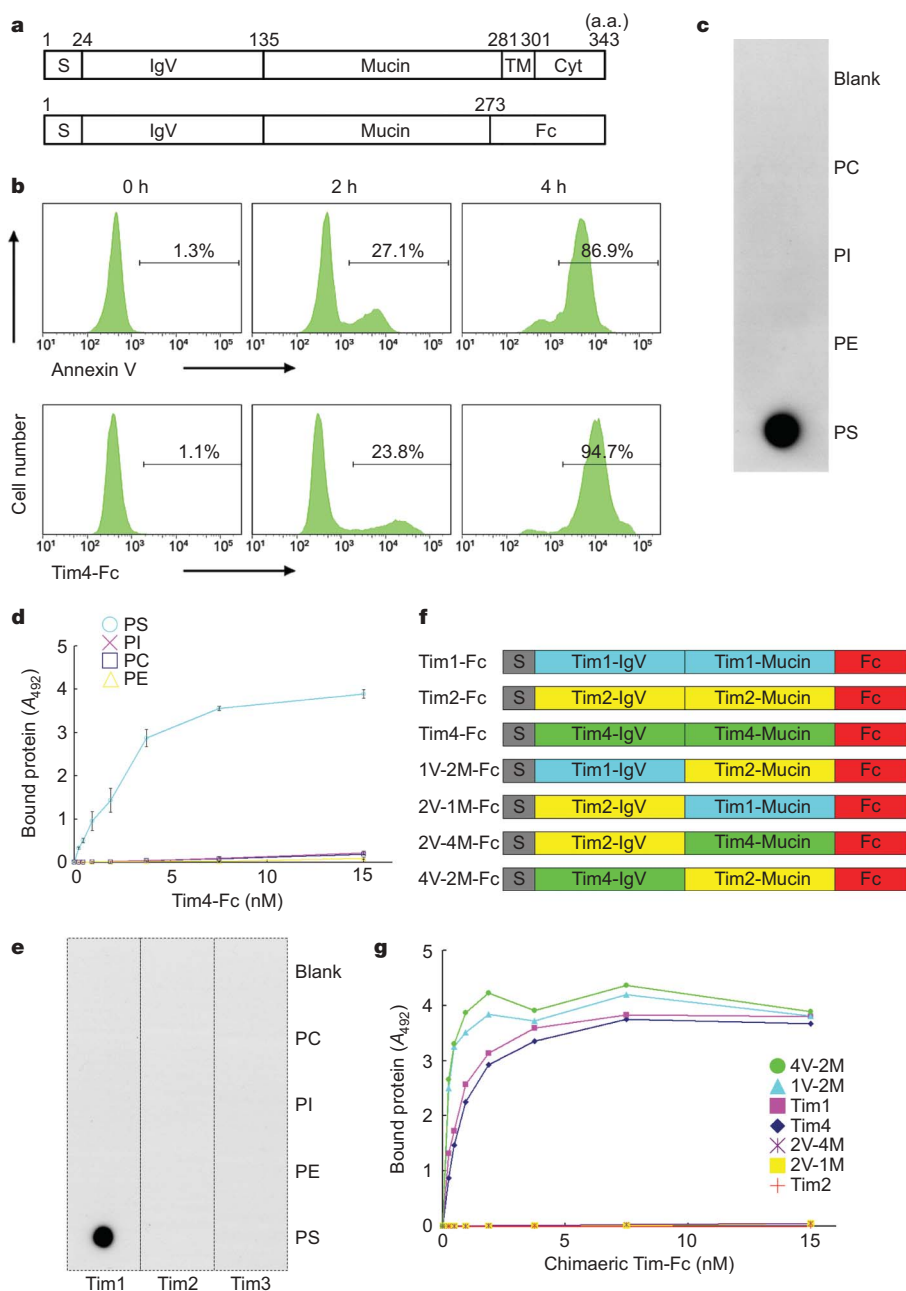
**Figure 2 | Tim4-mediated engulfment of apoptotic cells.** **a**, NIH3T3/DNase II (filled squares, open squares) and NIH3T3/DNase II/*Tim4* (filled circles, open circles) incubated with  $3.0 \times 10^5$  (filled squares, filled circles) or  $6.0 \times 10^5$  (open squares, open circles) apoptotic *CAD*<sup>-/-</sup> thymocytes for the indicated periods. The percentage of TUNEL<sup>+</sup> macrophages was determined by flow cytometry. Average values from three experiments are shown; error bars,  $\pm 1$  s.e. **b**, NIH3T3/DNase II and NIH3T3/DNase II/*Tim4* used for 60-min phagocytosis, stained for TUNEL, and observed by microscopy. Arrows indicate TUNEL<sup>+</sup> cells. **c**, NIH3T3/DNase II or NIH3T3/DNase II/*Tim4* incubated with healthy or apoptotic thymocytes in the presence of Kat5-18, stained for TUNEL. The percentage of TUNEL<sup>+</sup> macrophages was determined

by flow cytometry. Average values from three experiments are shown; error bars,  $\pm 1$  s.e. **d**, *CAD*<sup>-/-</sup> mice that had received normal hamster Ab or Kat5-18 were treated with dexamethasone. Top, the thymus stained for TUNEL (green) and F4/80 (red). The number of TUNEL<sup>+</sup> cells determined for 50 randomly selected macrophages (bottom). Experiments were carried out three times, and the percentage of the macrophages carrying more than 15 apoptotic cells plotted (right). Mann–Whitney’s U-test was used to analyse the difference, and *P* values are shown. **e**, Normal hamster IgG, Kat5-18 or PBS was injected into mice twice a week for 5 weeks. The serum concentration of anti-cardiolipin and anti-dsDNA was determined at 2, 5 and 10 weeks. *P* values determined as above are shown. *A*<sub>492</sub>, absorbance at 492 nm.

cells were cultured with NIH3T3/DNase II, few apoptotic cells were seen attached to NIH3T3/DNase II (Fig. 2b). In contrast, many apoptotic cells associated with NIH3T3/DNase II/Tim4, and many were TUNEL-positive, indicating that these cells had been engulfed. NIH3T3/DNase II/Tim4 engulfed apoptotic cells but not healthy ones, and this engulfment was dose-dependently inhibited by Kat5-18 (Fig. 2c), confirming that Tim4 conferred the ability to engulf apoptotic cells on NIH3T3/DNase II. To confirm the function of Tim4 *in vivo*, *CAD*<sup>-/-</sup> mice were given Kat5-18, and treated with dexamethasone to induce apoptosis in the thymus. As shown in Fig. 2d, in the thymus of mice that received the control IgG, 32.3% of F4/80<sup>+</sup> macrophages carried more than 15 TUNEL<sup>+</sup> apoptotic cells. This percentage of the macrophages carrying more than 15

apoptotic cells was reduced to about 12.6% in the thymus of the mice that received Kat5-18. Inefficient engulfment of apoptotic cells often leads to autoimmune diseases<sup>17</sup>. In fact, the mice that received Kat5-18 developed autoantibodies (anti-cardiolipin and anti-double-stranded-DNA antibodies) within 5 weeks in their serum (Fig. 2e). Two mice strongly responded to Kat5-18 and two weakly, but the injection of normal IgG or PBS never caused the development of autoantibodies. The serum anti-cardiolipin antibody level was reduced when the injection of Kat5-18 was interrupted, confirming that this increase was due to Kat5-18.

Mouse Tim4 is a type I membrane protein, consisting of a signal sequence and extracellular, transmembrane and cytosolic regions (Fig. 3a). To examine how Tim4 enhances the engulfment of



**Figure 3 | Binding to phosphatidylserine.** **a**, Structure of Tim4, showing signal sequence (S), IgV-like (IgV), mucin-like (Mucin), transmembrane (TM) and cytosolic regions (Cyt). a.a., amino acid. Tim4-Fc carries the Tim4 extracellular region and the Fc region of human IgG. **b**, Ba/F3-Fas treated with Fas ligand for the indicated periods, stained with annexin V or Tim4-Fc. **c**, A filter spotted with phosphatidylcholine (PC), phosphatidylinositol (PI), phosphatidylethanolamine (PE) or phosphatidylserine (PS), incubated with

Tim4-Fc followed by anti-IgG. **d**, Microtitre plates coated with indicated phospholipids incubated with Tim4-Fc, followed by anti-IgG. The assay was performed in triplicate. **e**, Lipid overlay assay performed as **c** with Tim1-Fc, Tim2-Fc and Tim3-Fc. **f**, Structures of the extracellular region of Tim1, Tim2 and Tim4, and their hybrids, fused to human IgG. **g**, Microtitre plates coated with PS incubated with increasing concentrations of Tim1-Fc, Tim2-Fc, Tim4-Fc, or their hybrids.



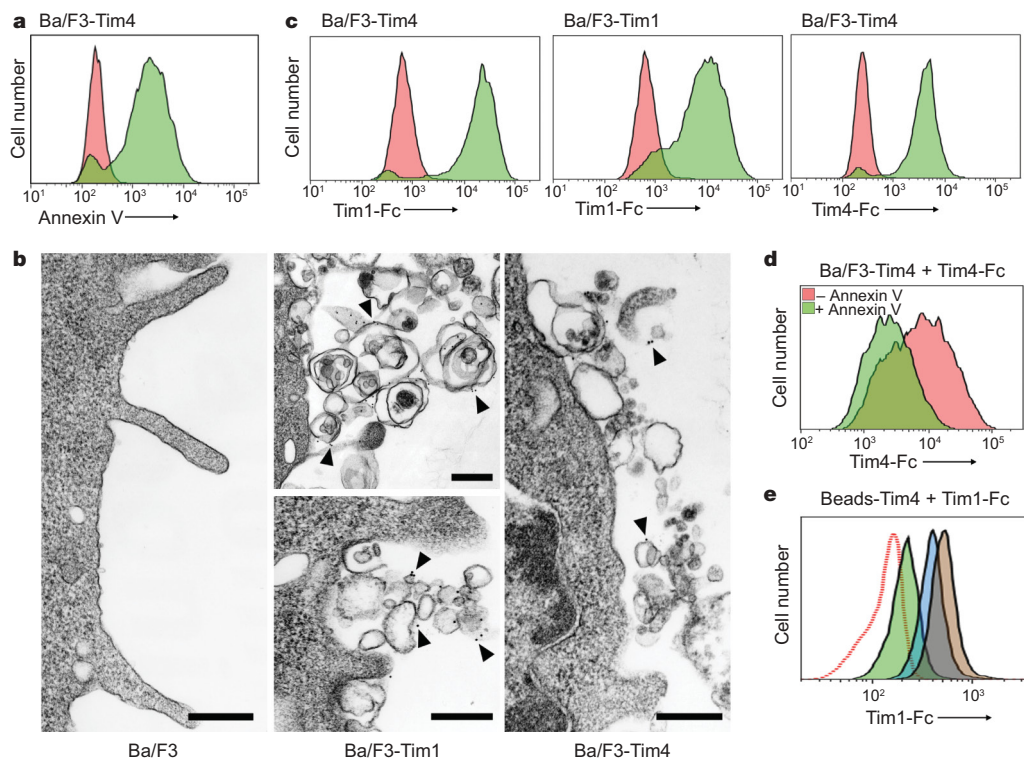
apoptotic cells, the Tim4 extracellular region was fused to the Fc region of human IgG (Tim4-Fc). The purified Tim4-Fc showed a band at 82 kDa in SDS-PAGE under reducing conditions, and a 170 kDa band under non-reducing conditions (Supplementary Fig. 3), indicating that Tim4-Fc existed as a dimer. When Ba/F3 expressing Fas was treated with Fas ligand, they underwent apoptosis and bound annexin V (Fig. 3b), which binds to PS exposed on the surface of apoptotic cells<sup>18</sup>. These cells also bound Tim4-Fc, with a time course similar to their binding of annexin V. When a nitrocellulose filter spotted with various phospholipids was incubated with Tim4-Fc, Tim4-Fc bound PS, but not phosphatidylcholine, phosphatidylinositol, or phosphatidylethanolamine (Fig. 3c). The affinity of Tim4-Fc for PS, determined with microtitre plates coated with phospholipids, was comparable to that of MFG-E8 ( $K_d \approx 2$  nM) (Fig. 3d). These results suggested that Tim4 recognizes PS exposed on apoptotic cells as an engulfment signal. In support of this idea, the MFG-E8 mutant D89E dose-dependently inhibited the engulfment of apoptotic cells by NIH3T3/DNase II/Tim4 (Supplementary Fig. 4).

The Tim family has three members in human (TIM1, 3 and 4), and eight in mouse (Tim1–8). To examine whether other Tim family members can recognize PS, the Fc fusion proteins were prepared with mouse Tim1, Tim2 and Tim3 (Supplementary Fig. 5). Tim1-Fc, Tim2-Fc and Tim3-Fc each behaved as a dimer (Supplementary Fig. 6). The lipid overlay assay indicated that Tim1-Fc, but neither Tim2-Fc nor Tim3-Fc, specifically bound PS (Fig. 3e). Accordingly, full-length Tim1, but neither Tim2 nor Tim3, enhanced the engulfment of apoptotic cells when it was expressed in NIH3T3/DNase II cells (Supplementary Fig. 7). The extracellular region of Tim family members consists of two domains (IgG V-chain like (IgV) and mucin domains)<sup>8</sup>. To examine which domain of Tim1 and Tim4 is responsible for binding PS, IgV and mucin domains were swapped between

Tim1 and Tim2, and between Tim4 and Tim2, and fused to the Fc region (Fig. 3f). As shown in Fig. 3g, the chimaeric molecules carrying IgV domain from Tim1 or Tim4 tightly bound PS, while no binding was observed with the chimaeric molecules carrying mucin domain from Tim1 or Tim4.

Various mammalian cells produce exosomes that expose PS on their surface<sup>7</sup>. The ability of Tim1 and Tim4 to bind PS suggested that exosomes bind to cells expressing Tim1 or Tim4. To examine this possibility, Ba/F3 cell lines expressing Tim1 or Tim4 (Ba/F3-Tim1 or Ba/F3-Tim4) but not the parental Ba/F3 bound annexin V (Fig. 4a, and data not shown). Ba/F3-Tim1 and Ba/F3-Tim4 were then treated with biotin-labelled annexin V, followed by staining with gold-conjugated streptavidin. Observation by electron microscopy indicated that Ba/F3-Tim1 or Ba/F3-Tim4 was associated with exosome-like vesicles, and several gold particles were found on the outer leaflet of the exosomes (Fig. 4b). Such vesicles were not observed with the Ba/F3 parental cells. Meyers *et al.*<sup>19</sup> reported that Tim4 acts as a ligand for Tim1 expressed in Th2 cells. Accordingly, Tim1-Fc strongly bound Ba/F3-Tim4 (Fig. 4c). In addition, Tim1-Fc bound Ba/F3-Tim1, and Tim4-Fc bound Ba/F3-Tim4. The ability of Tim4 and Tim1 to bind PS suggested that these homophilic and heterophilic interactions between Tim1 and Tim4 can be stimulated by exosomes. In support of this idea, the binding of Tim4-Fc to Ba/F3-Tim4 was inhibited by annexin V (Fig. 4d). To directly confirm the effect of exosomes on association between Tim1 and Tim4, microspheres were coated with Tim4-Fc, and subjected to binding assay with Tim1-Fc. As shown in Fig. 4e, little Tim1-Fc bound to Tim4-Fc-coated beads, and exosomes from Ba/F3 dose-dependently stimulated the binding of Tim1-Fc to the beads.

Phosphatidylserine exposed on the outer leaflet of the plasma membrane is often used as a recognition signal for phagocytes to



**Figure 4 | Association of Tim1 and Tim4 via exosomes.** **a**, Ba/F3-Tim4 incubated with PE-annexin V, and analysed by flow cytometry. Staining profile of Ba/F3 in red. **b**, Electron micrographs of Ba/F3, Ba/F3-Tim1 or Ba/F3-Tim4, incubated with biotin-annexin V and stained with streptavidin-colloidal gold. Gold particles on exosome-like vesicles indicated by arrowheads. Scale bar, 0.3  $\mu$ m. **c**, Ba/F3-Tim4 or Ba/F3-Tim1 incubated with Tim1-Fc or Tim4-Fc, stained with FITC-anti-IgG, and analysed by flow

cytometry. Staining profiles of Ba/F3 in red. **d**, Ba/F3-Tim4 incubated with Tim4-Fc in the absence or presence of annexin V, stained with FITC-anti-IgG. **e**, Tim4-Fc-coated latex beads incubated with Tim1-Fc in the absence (green) or presence of 10-fold (blue) or 50-fold (brown) concentrated exosomes, stained with PE-anti-Tim1, and analysed by flow cytometry. Profile without Tim1-Fc in red dotted line.

engulf apoptotic cells. Fadok *et al.* reported that phosphatidylserine receptor (PSR), a ubiquitously expressed type II-membrane protein, binds to PS on apoptotic cells and mediates their engulfment<sup>20</sup>, and initial experiments with *PSR*<sup>-/-</sup> mice supported this notion<sup>21,22</sup>. However, a careful analysis of *PSR*<sup>-/-</sup> mice by Bose *et al.*<sup>23</sup> indicated that PSR is not involved in the engulfment of apoptotic cells, and it is unlikely to work as a receptor for PS. Here, we showed that Tim4 expressed in macrophages in various tissues mediates the engulfment of apoptotic cells by recognizing PS, indicating that Tim4 is the phosphatidylserine receptor for engulfment of apoptotic cells. Tim1 also enhanced the PS-dependent engulfment of apoptotic cells. Tim1 is expressed in kidney tubule cells post-ischaemia<sup>24</sup>, suggesting that it may have a role in removing injured and dead cells, to help restore the morphological integrity of the kidney. It will be interesting to examine whether any other molecules work as phosphatidylserine receptors.

Some non-apoptotic T or B cells<sup>25,26</sup> as well as macrophages<sup>27</sup> have been reported to bind annexin V. Here, we showed that the cells expressing Tim1 or Tim4 associated with exosomes that carried exposed PS. Activated Th2 cells and B cells express Tim1<sup>28,29</sup>, while the resident macrophages express Tim4, suggesting that the non-apoptotic annexin V-binding cells are those that express Tim1 or Tim4. Exosomes are secreted from various cells and act as an inter-cellular signalling device<sup>7</sup>. Tim1 and Tim4 may mediate this signalling by enhancing the uptake of exosomes. Cells expressing Tim1 or Tim4 tended to aggregate (M.M. and S.N., unpublished results). PS on activated T cells is suggested to function in cell-to-cell contact at the immunological synapse<sup>25</sup>. Thus, Tim1 and Tim4 may also be involved in cell-to-cell interactions via PS on exosomes. Finally, Tim family genes of human and mouse are located on a genetic interval that has linkage to a number of autoimmune diseases, such as asthma, allergy and atopy<sup>30</sup>. The establishment of Tim1 and Tim4 as PS receptors may contribute to the understanding of these autoimmune diseases.

## METHODS SUMMARY

**Production of monoclonal antibodies.** Armenian hamsters (Oriental Yeast) were immunized with mouse peritoneal cells, and about 1,200 hybridomas were established by fusing cells from the popliteal and inguinal lymph nodes with NSO<sup>bcl-2</sup> mouse myeloma. The hybridoma supernatants were tested by a phagocytosis assay, and one hybridoma (Kat5-18) that inhibited the engulfment of apoptotic cells was identified.

**Construction of cDNA library, and expression cloning.** The double-stranded cDNA longer than 1.0 kb was ligated into a *Bst*XI-digested pMXs vector<sup>15</sup>. *Escherichia coli* DH10B was transformed by electroporation, and about  $1.0 \times 10^6$  clones were produced. Plasmid DNA from the cDNA library was introduced into PLAT-E packaging cells, and the culture supernatant was used to infect Ba/F3. A subpopulation of Ba/F3 that stained intensely with Kat5-18 was selected by repeated FACS. The integrated cDNA was recovered by PCR with the vector primers from the genomic DNA of Ba/F3 transformants.

**Phagocytosis assay.** To assay the engulfment of apoptotic cells, *CAD*<sup>-/-</sup> thymocytes were induced to undergo apoptosis with leucine-zipper tagged Fas ligand, and co-cultured with peritoneal macrophages or NIH3T3/DNase II cells. The cells were fixed with paraformaldehyde, subjected to TUNEL staining, and analysed by flow cytometry using a FACSaria (BD Biosciences).

Received 10 September; accepted 26 September 2007.

Published online 24 October 2007.

1. Danial, N. N. & Korsmeyer, S. J. Cell death: critical control points. *Cell* **116**, 205–219 (2004).
2. Savill, J., Dransfield, I., Gregory, C. & Haslett, C. A blast from the past: clearance of apoptotic cells regulates immune responses. *Nature Rev. Immunol.* **2**, 965–975 (2002).
3. Fadok, V. A. *et al.* Exposure of phosphatidylserine on the surface of apoptotic lymphocytes triggers specific recognition and removal by macrophages. *J. Immunol.* **148**, 2207–2216 (1992).

4. Yoshida, H. *et al.* Phosphatidylserine-dependent engulfment by macrophages of nuclei from erythroid precursor cells. *Nature* **437**, 754–758 (2005).
5. Tanaka, Y. & Schroit, A. J. Insertion of fluorescent phosphatidylserine into the plasma membrane of red blood cells. Recognition by autologous macrophages. *J. Biol. Chem.* **258**, 11335–11343 (1983).
6. Schlegel, R. A. & Williamson, P. Phosphatidylserine, a death knell. *Cell Death Differ.* **8**, 551–563 (2001).
7. Thery, C., Zitvogel, L. & Amigorena, S. Exosomes: composition, biogenesis and function. *Nature Rev. Immunol.* **2**, 569–579 (2002).
8. Kuchroo, V. K., Umetsu, D. T., DeKruyff, R. H. & Freeman, G. J. The TIM gene family: emerging roles in immunity and disease. *Nature Rev. Immunol.* **3**, 454–462 (2003).
9. Nagata, S. DNA degradation in development and programmed cell death. *Annu. Rev. Immunol.* **23**, 853–875 (2005).
10. Hanayama, R. *et al.* Identification of a factor that links apoptotic cells to phagocytes. *Nature* **417**, 182–187 (2002).
11. Hanayama, R. & Nagata, S. Impaired involution of mammary glands in the absence of milk fat globule EGF factor 8. *Proc. Natl Acad. Sci. USA* **102**, 16886–16891 (2005).
12. Hanayama, R. *et al.* Autoimmune disease and impaired uptake of apoptotic cells in MFG-E8-deficient mice. *Science* **304**, 1147–1150 (2004).
13. Miyasaka, K., Hanayama, R., Tanaka, M. & Nagata, S. Expression of milk fat globule epidermal growth factor 8 in immature dendritic cells for engulfment of apoptotic cells. *Eur. J. Immunol.* **34**, 1414–1422 (2004).
14. Hu, B., Sonstein, J., Christensen, P. J., Punturieri, A. & Curtis, J. L. Deficient *in vitro* and *in vivo* phagocytosis of apoptotic T cells by resident murine alveolar macrophages. *J. Immunol.* **165**, 2124–2133 (2000).
15. Kitamura, T. *et al.* Retrovirus-mediated gene transfer and expression cloning: powerful tools in functional genomics. *Exp. Hematol.* **31**, 1007–1014 (2003).
16. Kawane, K. *et al.* Impaired thymic development in mouse embryos deficient in apoptotic DNA degradation. *Nature Immunol.* **4**, 138–144 (2003).
17. Gaip, U. S. *et al.* Impaired clearance of dying cells in systemic lupus erythematosus. *Autoimmun. Rev.* **4**, 189–194 (2005).
18. Koopman, G. *et al.* Annexin V for flow cytometric detection of phosphatidylserine expression on B cells undergoing apoptosis. *Blood* **84**, 1415–1420 (1994).
19. Meyers, J. H. *et al.* TIM-4 is the ligand for TIM-1, and the TIM-1-TIM-4 interaction regulates T cell proliferation. *Nature Immunol.* **6**, 455–464 (2005).
20. Fadok, V. A. *et al.* A receptor for phosphatidylserine-specific clearance of apoptotic cells. *Nature* **405**, 85–90 (2000).
21. Li, M. O., Sarkisian, M. R., Mehal, W. Z., Rakic, P. & Flavell, R. A. Phosphatidylserine receptor is required for clearance of apoptotic cells. *Science* **302**, 1560–1563 (2003).
22. Kunisaki, Y. *et al.* Defective fetal liver erythropoiesis and T lymphopoiesis in mice lacking the phosphatidylserine receptor. *Blood* **103**, 3362–3364 (2004).
23. Bose, J. *et al.* The phosphatidylserine receptor has essential functions during embryogenesis but not in apoptotic cell removal. *J. Biol.* **3**, 15 (2004).
24. Ichimura, T. *et al.* Kidney injury molecule-1 (KIM-1), a putative epithelial cell adhesion molecule containing a novel immunoglobulin domain, is up-regulated in renal cells after injury. *J. Biol. Chem.* **273**, 4135–4142 (1998).
25. Fischer, K. *et al.* Antigen recognition induces phosphatidylserine exposure on the cell surface of human CD8<sup>+</sup> T cells. *Blood* **108**, 4094–4101 (2006).
26. Dillon, S. R., Mancini, M., Rosen, A. & Schlissel, M. S. Annexin V binds to viable B cells and colocalizes with a marker of lipid rafts upon B cell receptor activation. *J. Immunol.* **164**, 1322–1332 (2000).
27. Callahan, M. K., Williamson, P. & Schlegel, R. A. Surface expression of phosphatidylserine on macrophages is required for phagocytosis of apoptotic thymocytes. *Cell Death Differ.* **7**, 645–653 (2000).
28. Umetsu, S. E. *et al.* TIM-1 induces T cell activation and inhibits the development of peripheral tolerance. *Nature Immunol.* **6**, 447–454 (2005).
29. Gielen, A. W. *et al.* Expression of T cell immunoglobulin- and mucin-domain-containing molecules-1 and -3 (TIM-1 and -3) in the rat nervous and immune systems. *J. Neuroimmunol.* **164**, 93–104 (2005).
30. McIntire, J. J. *et al.* Identification of Tapr (an airway hyperreactivity regulatory locus) and the linked Tim gene family. *Nature Immunol.* **2**, 1109–1116 (2001).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank R. Hanayama for advice in the initial stage of this project, and M. Fujii and M. Harayama for secretarial assistance. This work was supported in part by Grants-in-Aid from the Ministry of Education, Science, Sports, and Culture in Japan.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to S.N. (snagata@four.med.kyoto-u.ac.jp).

## LETTERS

# SIRT1 regulates the histone methyl-transferase SUV39H1 during heterochromatin formation

Alejandro Vaquero<sup>1,†</sup>, Michael Scher<sup>1,3</sup>, Hediye Erdjument-Bromage<sup>4</sup>, Paul Tempst<sup>4</sup>, Lourdes Serrano<sup>2</sup> & Danny Reinberg<sup>1,3</sup>

In contrast to stably repressive, constitutive heterochromatin and stably active, euchromatin, facultative heterochromatin has the capacity to alternate between repressive and activated states of transcription<sup>1</sup>. As such, it is an instructive source to understand the molecular basis for changes in chromatin structure that correlate with transcriptional status. Sirtuin 1 (SIRT1) and suppressor of variegation 3–9 homologue 1 (SUV39H1) are amongst the enzymes responsible for chromatin modulations associated with facultative heterochromatin formation. SUV39H1 is the principal enzyme responsible for the accumulation of histone H3 containing a trimethyl group at its lysine 9 position (H3K9me3) in regions of heterochromatin<sup>2</sup>. SIRT1 is an NAD<sup>+</sup>-dependent deacetylase that targets histone H4 at lysine 16 (refs 3 and 4), and through an unknown mechanism facilitates increased levels of H3K9me3 (ref. 3). Here we show that the mammalian histone methyltransferase SUV39H1 is itself targeted by the histone deacetylase SIRT1 and that SUV39H1 activity is regulated by acetylation at lysine residue 266 in its catalytic SET domain. SIRT1 interacts directly with, recruits and deacetylates SUV39H1, and these activities independently contribute to elevated levels of SUV39H1 activity resulting in increased levels of the H3K9me3 modification. Loss of SIRT1 greatly affects SUV39H1-dependent H3K9me3 and impairs localization of heterochromatin protein 1. These findings demonstrate a functional link between the heterochromatin-related histone methyltransferase SUV39H1 and the histone deacetylase SIRT1.

SIRT1 is a member of the Sir2 family of NAD<sup>+</sup>-dependent histone deacetylases<sup>3,4</sup> and promotes heterochromatin formation through the coordination of several events<sup>3</sup>. SIRT1 deacetylates histones (H4K16 and H3K9), recruits histone H1b, and promotes the loss of a mark associated with transcriptionally active chromatin, H3K79me2, and the establishment of marks associated with repressed chromatin such as H3K9me3 and H4K20me1 (ref. 3). How SIRT1 affects the levels of histone modifications other than deacetylation remains unclear.

We first tested whether SIRT1 might recruit an enzyme responsible for H3K9me3, during purification of Flag-tagged SIRT1 expressed in human 293 cells. Indeed, immunoprecipitated SIRT1 contained histone H3 lysine methyltransferase (HKMT) activity (Fig. 1a, lanes 5–7), the levels of which were greatly decreased in a similarly purified but catalytically inactive form of SIRT1 (lanes 8–10)<sup>3</sup>. This HKMT activity was specific for H3K9 (Fig. 1b). Histone octamers reconstituted with recombinant histones containing either wild-type or mutant H3, in which Lys 27 was replaced with Ala (K27A), exhibited similar levels of methylation. In contrast, mutant H3, in which Lys 9 was replaced with Ala (K9A), was an inappropriate substrate.

Two candidates for this specific HKMT were SUV39H1 and the euchromatic histone-lysine N-methyltransferase 2 G9A (also known as EHMT2), given their specificity for H3K9me formation<sup>5</sup>. Interaction between SIRT1 and G9A was undetectable (not shown). However, when 293 cells were co-transfected with Flag–SIRT1 and a Myc-tagged version of the major cellular activity for H3K9me3 (SUV39H1), the presence of either resulted in the co-immunoprecipitation of the other in pull-down experiments (Fig. 1c, d). The interaction between SUV39H1 and SIRT1 was specific, because the related NAD<sup>+</sup>-dependent deacetylase SIRT2 failed to pull-down SUV39H1 (Fig. 1c). This is consistent with the functional link found between the homologues of SUV39H1 and SIRT1 in *Schizosaccharomyces pombe* (Clr4 and Sir2, respectively)<sup>6</sup>. Chromatin immunoprecipitation experiments were then performed on cells containing an inducible SIRT1 fused to the yeast Gal4 DNA-binding protein domains<sup>3</sup> that were transfected with Myc-tagged SUV39H1 (Fig. 1e). Myc–SUV39H1 was present at the integrated thymidine kinase promoter containing Gal4 sites in these cells only on induction of Gal4–SIRT1 with which it colocalized, demonstrating their interaction *in vivo*. This correlated with a significant loss in H3K9ac levels and enrichment in H3K9me3 levels at the promoter.

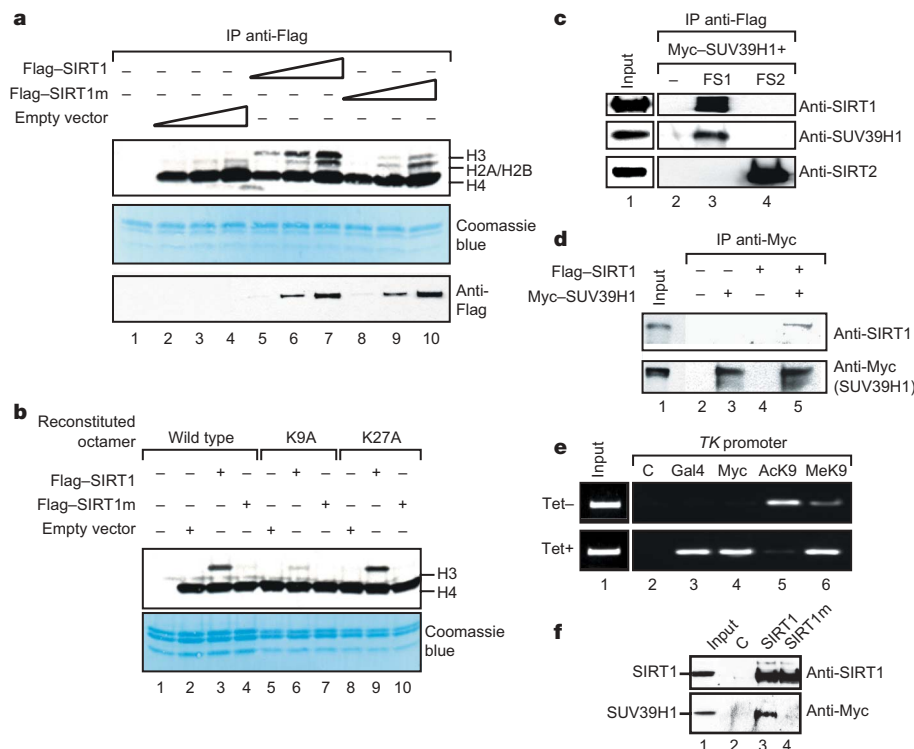
Co-immunoprecipitation experiments were performed to identify the interaction domains. A point mutation that renders SIRT1 catalytically inactive caused deficiency in interaction (Fig. 1f). The amino terminus of SIRT1 that interacts with histone H1b (ref. 3) also interacts with SUV39H1 (see below). Removal of either the SET domain or the chromo domain of SUV39H1 decreased its interaction with SIRT1, whereas removal of the first 89 N-terminal residues including the chromo domain eliminated interaction (Supplementary Fig. 1a). The chromo domain alone exhibited detectable interaction (Supplementary Fig. 1b), indicating that the N-terminal domain (1–43) and the chromo domain (44–88) of SUV39H1 were both required for interaction with SIRT1. This is similar to the case reported for SUV39H1 interaction with the polycomb human protein PC2 (ref. 7).

To understand how SIRT1 affects global levels of H3K9me3, we tested for its effect on the levels of SUV39H1-specific activity using methyltransferase assays performed *in vitro* with histone octamers as substrates. Increasing amounts of recombinant SIRT1 led to elevated levels of SUV39H1-mediated methylation of histone H3 (Fig. 2a, compare lane 2 with lanes 4–6), whereas recombinant SIRT2 (Fig. 2a, lanes 8–10) and bovine serum albumin (not shown) were ineffectual. The increased SUV39H1 activity exhibited the expected substrate activity, being unable to methylate recombinant octamers containing mutant H3K9A or mononucleosomes (Supplementary Fig. 2). The SIRT1-mediated stimulation of SUV39H1 activity was

<sup>1</sup>Howard Hughes Medical Institute, Division of Nucleic Acids Enzymology, Department of Biochemistry, University of Medicine and Dentistry of New Jersey, Robert Wood Johnson Medical School, New Jersey 08854, USA. <sup>2</sup>Department of Genetics, Human Genetics Institute, Rutgers University, 145 Bevier Road, Piscataway, New Jersey 08854, USA.

<sup>3</sup>Department of Biochemistry, NYU-Medical School, 522 First Avenue, New York, New York 10016, USA. <sup>4</sup>Molecular Biology Program, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA. <sup>†</sup>Present address: ICREA and IMBB-CSIC/IRB, Parc Científic de Barcelona, Josep Samitier 1–5, 08028 Barcelona, Spain.

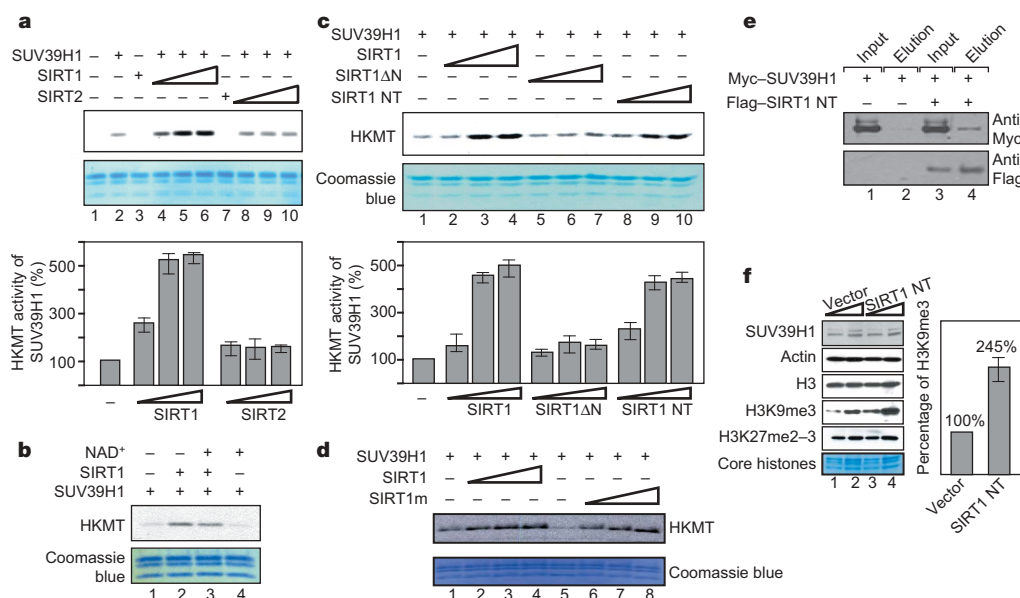




**Figure 1 | The histone lysine methyltransferase SUV39H1 interacts with catalytically active SIRT1 *in vitro* and *in vivo*.** **a**, Affinity-purified Flag-tagged SIRT1 was assayed for the presence of histone methyltransferase activity (top). Recombinant histone octamers used as substrates were stained with Coomassie blue (middle). Flag-tagged wild-type and catalytically inactive SIRT1 (H363Y)<sup>3</sup> were analysed in western blots (bottom). SIRT1m, SIRT1 mutant. **b**, The HKMT is specific for histone H3 lysine 9 in HKMT assays. Assays were performed as in **a**, but using octamers reconstituted with histones as indicated. **c**, **d**, SIRT1 and SUV39H1 exhibit reciprocal co-immunoprecipitation *in vitro*, as shown in western blots. FS1, Flag-tagged SIRT1; FS2, Flag-tagged SIRT2. **e**, Chromatin immunoprecipitation using cells carrying tetracycline (Tet)-inducible Gal4-SIRT1 and an integrated luciferase reporter, transfected with Myc-tagged SUV39H1. The generation of these cells was described earlier<sup>3</sup>. The probe is specific for the promoter of the integrated thymidine kinase (TK) reporter containing Gal4 sites. C, control. **f**, Western blot of anti-Flag antibody co-immunoprecipitations performed *in vitro* with Myc-SUV39H1 and control (C), Flag-SIRT1 (SIRT1) or Flag-SIRT1 mutant (SIRT1m)<sup>3</sup>.

NAD<sup>+</sup>-independent (Fig. 2b), but the N terminus of SIRT1 (Fig. 2c) was required, as it was for SUV39H1 interaction (Fig. 2e). Titration of recombinant protein consisting of only the N-terminal 250 residues of SIRT1 was sufficient to stimulate the methyltransferase activity (Fig. 2c), and a baculovirus-expressed catalytically inactive

mutant SIRT1, which contains an intact N-terminal domain, was similarly effective (Fig. 2d). In addition, overexpression of the SIRT1 N terminus in 293 cells gave rise to elevated levels of H3K9me3 *in vivo* without affecting the levels of SUV39H1, histone H3, or di- or tri-methylated lysine residue 27 of histone H3 (Fig. 2f).



**Figure 2 | SIRT1 upregulates SUV39H1 activity *in vitro* and *in vivo* through the SIRT1 N terminus.** **a**, Methyltransferase (HKMT) activity of recombinant SUV39H1 as a function of increasing amounts of recombinant baculovirus-expressed SIRT1 or bacterially expressed SIRT2. Endogenous histone octamers (purified from HeLa cells) used as substrates were stained with Coomassie blue. Quantifications of multiple experiments are represented below and compared to the HKMT activity of SUV39H1 alone (lane 2), which is considered to be 100%. **b**, As in **a**, but in the absence or presence of NAD<sup>+</sup>. **c**, As in **a**, but with increasing amounts of either wild type or SIRT1 deletion mutants (expressed in *Escherichia coli*) that either do not contain the N terminus ( $\Delta$ N) or only contain the N terminus (NT). **d**, As in

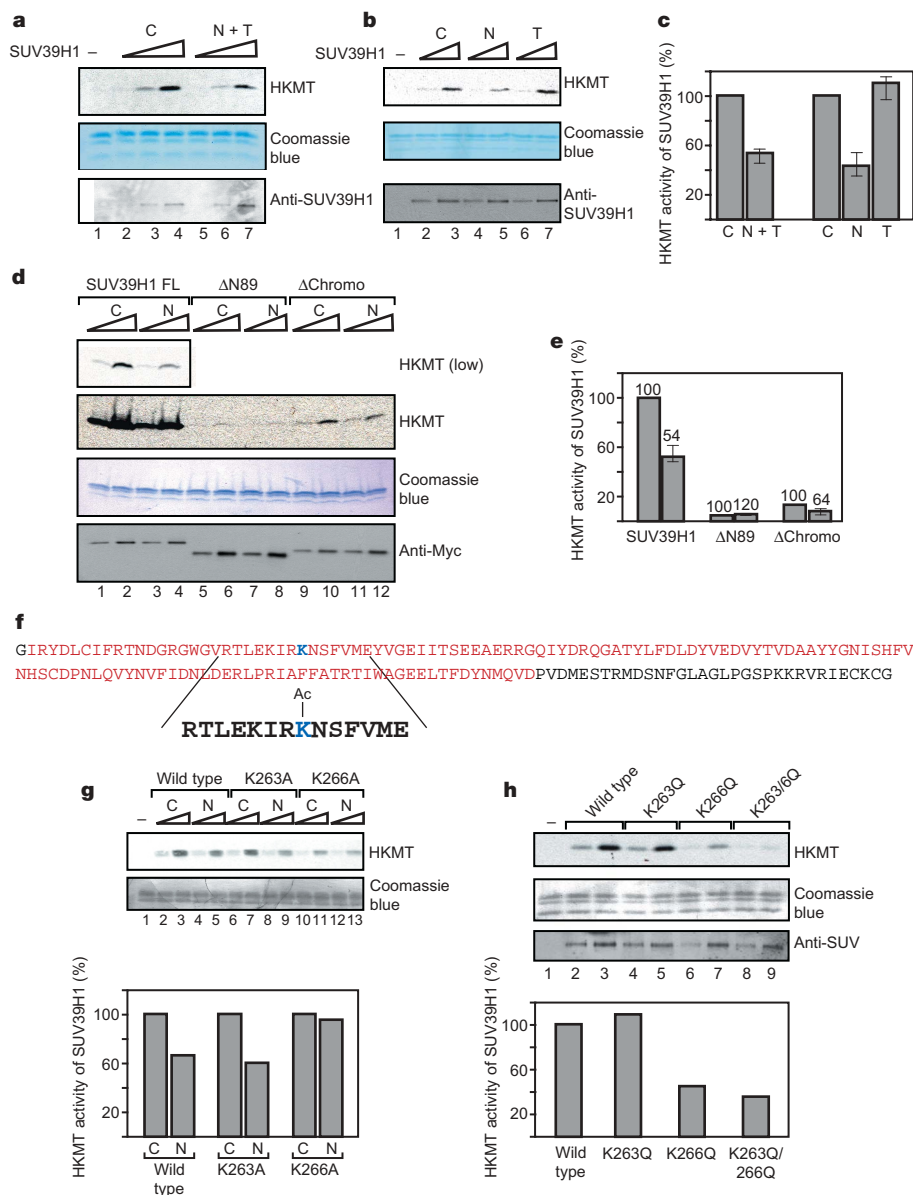
**a**, but with increasing amounts of either wild-type SIRT1 or the catalytically inactive point mutant H363Y (expressed in baculovirus). **e**, Western blots from immunoprecipitations as in Fig. 1c, but using the Flag-tagged SIRT1 N-terminal region (SIRT1 NT)<sup>3</sup> and Myc-SUV39H1 as indicated. **f**, Western blots performed on extracts from 293 cells transfected with either empty vector or vector expressing the SIRT1 N terminus using the specific antibodies indicated. Core histones are stained with Coomassie blue. The right side shows quantification of the increased levels of H3K9me3 on overexpression of the N-terminal 250 residues of SIRT1 (SIRT1 NT) relative to the vector control. Error bars represent s.d.  $n = 4$  for **a** and **c**;  $n = 3$  for **f**.

Thus, the N terminus of SIRT1 is required both for interaction with SUV39H1 and for stimulation of its activity. The conformation of SUV39H1 is probably altered in a manner favourable to its activity on interaction with the N terminus of SIRT1.

SIRT1 is the principal NAD<sup>+</sup>-dependent deacetylase in mammalian cells and, although SIRT1 targets histones, it also deacetylates a wide range of proteins<sup>3,8</sup>; therefore, we examined if SIRT1 could deacetylate SUV39H1. Human 293 cells overexpressing Myc-tagged SUV39H1 were exposed to the histone deacetylase (HDAC) inhibitor trichostatin A (TSA) (inhibitor of class I and II HDACs)<sup>9</sup> and to nicotinamide (inhibitor of the Sir2 family or class III HDACs)<sup>10</sup>, together or independently. SUV39H1 was then affinity-purified, its protein levels normalized by western blot, and was assayed for

HKMT activity. A clear loss in Suv39h1 activity was observed with both inhibitors (Fig. 3a, c), and this affect was specific to nicotinamide treatment (Fig. 3b, c). The absence of the chromo domain of SUV39H1 was ineffectual (Fig. 3d, e). Of note, the absence of its N terminus rendered SUV39H1 nearly inactive (Fig. 3d, e). That nicotinamide treatment decreased SUV39H1 activity *in vivo* strongly implicated the deacetylase activity of SIRT1. We next tested if this HKMT exists in an acetylated form in the cell and whether or not such acetylation might modulate its activity.

SUV39H1 derived from nicotinamide-treated cells exhibited acetylation that mapped to Lys 266 (Fig. 3f and Supplementary Fig. 3), as compared to untreated cells. This residue is present within the catalytic SET (Su(var)3-9, Enhancer-of-zeste, Trithorax) domains of



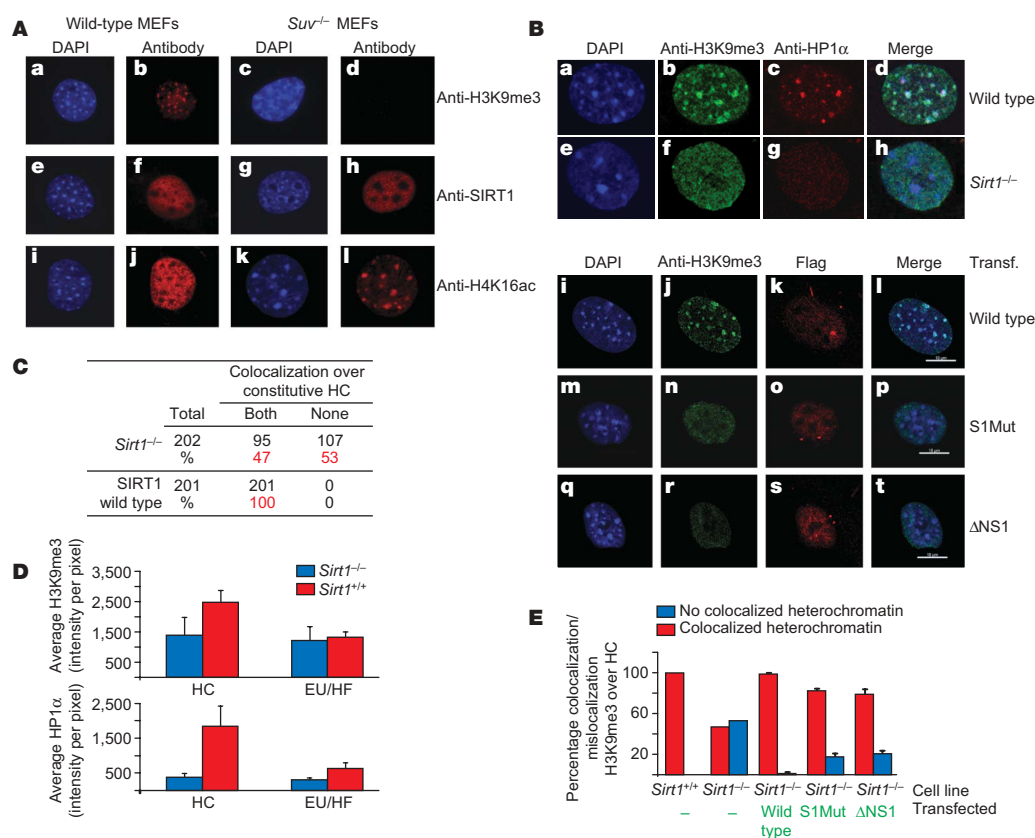
**Figure 3 | Acetylation of SUV39H1 negatively regulates its activity and is elevated in nicotinamide-treated cells.** **a, b**, SUV39H1 purified from 293 cells untreated or treated with inhibitors for histone deacetylases TSA (T) and nicotinamide (N) either together (**a**) or independently (**b**) and assayed for HKMT activity (upper panel). Histone proteins were stained with Coomassie blue (middle panel) and the levels of purified SUV39H1 were normalized with western blots using antibody specific to the Myc-tag (bottom panel). **c**, Quantification of results shown in **a** and **b**. **d**, Full length (FL) or indicated deletion mutants of Myc-tagged SUV39H1 purified after transfection of untreated (C) or nicotinamide (N)-treated 293 cells and

assayed as in **a** and **b** but with anti-Myc-specific antibody. **e**, Quantification of results shown in **d**. **f**, Amino acid sequence of the SET domain of SUV39H1 with Lys 266, which is subject to acetylation as indicated. **g, h**, Human 293 cells transfected with either wild type or mutants of Myc-SUV39H1 containing alanine substitutions (**g**) or glutamine substitutions (**h**) at the residues indicated. Cells were untreated (C) or treated with nicotinamide (N) and the proteins were purified with Myc-resin, normalized as in **a, b** and **d**, and assayed for HKMT activity as in **a, b** and **d**. In both **g** and **h**, quantifications of the HKMT assays are shown below. Error bars represent s.d. *n* = 4 for **c** and **e**.

EZH2, SUV420H1, SUV420H2 and SUV39H1, but not within that of yeast Clr4 (Supplementary Fig. 5a). SUV39H1 in which Lys 266 was replaced with Ala was unresponsive to nicotinamide treatment, whereas a control Ala substitution at Lys 263 behaved similarly to the wild type (Fig. 3g). In addition, a glutamine substitution that mimics an acetylated lysine residue was ineffectual at the Lys 263 position, but resulted in significant loss in SUV39H1 activity when at Lys 266 (Fig. 3h). This loss in activity reflected that observed in the case of Myc-tagged SUV39H1 derived from nicotinamide-treated cells (Fig. 3b). Additionally, SIRT1 can partially deacetylate purified SUV39H1 with lysine residues radiolabelled with acetic anhydride (Supplementary Fig. 4). This was specific to SIRT1, because neither SIRT2 nor SIRT3 deacetylated SUV39H1 although both were active in deacetylating H4K16ac and in the nicotinamide exchange reaction (data not shown and Supplementary Fig. 4). These results established that the enzymatic activity of SUV39H1 is negatively regulated by acetylation at residue 266, and that SIRT1-mediated deacetylation relieves this inhibition. Interestingly, comparison between the SET domain of SUV39H1 and the published structure of the SET

domains of two homologues, *Schizosaccharomyces pombe* Clr4 and *Neurospora crassa* Dim-5, indicates that Lys 266 is located in an exposed loop that participates in mediating the interaction between the SET domain and the cysteine-rich post-SET domain (Supplementary Fig. 5b). Acetylation of Lys 266 could alter this inter-domain interaction and affect the specific activity of the enzyme, which would explain the stimulatory effect of Lys 266 deacetylation on SUV39H1 activity<sup>11</sup>.

To corroborate this functional relationship, 293 cells were tested for colocalization of SIRT1 and green-fluorescent-protein-tagged SUV39H1. Consistent with published results<sup>2,12</sup>, SUV39H1 was found in constitutive heterochromatin foci, although it was also spread throughout the nucleus at distinctive regions (data not shown, Fig. 4A). Because the loss of SIRT1 results in decreased levels of H3K9me3 (ref. 3) and SIRT1 interacts with and activates SUV39H1 (as shown in this report), SIRT1 might also be affected when SUV39H1 is deficient. Immunofluorescence studies performed on mouse embryonic fibroblasts (MEFs, wild type or *Suv*<sup>-/-</sup>) demonstrated that SIRT1 localization was similar, although its



**Figure 4 | SIRT1 and SUV39H1 activities are linked in vivo.** **A**, Comparison of wild-type and *Suv*<sup>-/-</sup> (*Suv39h1/2*<sup>-/-</sup>) MEFs using immunofluorescence microscopy for the distribution of H3K9me3 (panels **b**, **d**), H4K16ac (**j**, **l**) and SIRT1 (**f**, **h**). DAPI staining of the cells is shown as a control for DNA and chromatin distribution (**a**, **c**, **e**, **g**, **i** and **k**). **B**, Comparison of wild-type and *Sirt1*<sup>-/-</sup> MEFs using immunofluorescence microscopy for the distribution of DAPI (**a**, **e**), H3K9me3 (**b**, **f**) and HP1α (**c**, **g**). A merged image of them is shown (**d**, **h**). *Sirt1*<sup>-/-</sup> MEFs were transfected with Flag-tagged SIRT1 (wild type), the SIRT1 catalytically inactive point mutant H363Y (S1Mut) or ΔNSIRT1 (ΔNS1), and were analysed by immunofluorescence for the distribution of H3K9me3 (**j**, **n**, **r**), Flag tag (**k**, **o**, **s**) and DAPI staining (**i**, **m**, **q**). A merged image of DAPI and H3K9me3 is shown (**l**, **p**, **t**). The images shown are representative of the experiment. Full quantification of the data (percentage of cells that show H3K9me3 localization in heterochromatin on each condition) is included as a graph in **E**. **C**, Quantification of cells in **A**. Significant numbers of the cells (202 for *Sirt1*<sup>-/-</sup> and 201 for wild type) were checked for colocalization of HP1α and H3K9me3 over DAPI bright blue regions. Cells showing colocalization were

counted and expressed as a frequency of the total number of cells. Data were obtained from two different slides per group. **D**, Histograms show the average intensity per pixel of H3K9me3 (upper panel) and HP1α (lower panel) within constitutive heterochromatin (HC) and outside of constitutive heterochromatin, that is, euchromatin (EU) and facultative heterochromatin (HF), for *Sirt1*<sup>-/-</sup> and wild-type MEFs. The average HP1 and H3K9me3 intensity per pixel is higher in *Sirt1*<sup>+/+</sup> than *Sirt1*<sup>-/-</sup> within constitutive heterochromatin regions ( $P < 0.001$ ). Error bars represent s.d. **E**, Quantification of the rescue experiments shown in **B**, panels **i**–**t**, represented in a graph showing the percentage of colocalization of H3K9me3 with heterochromatin (red bars), and the corresponding percentage of mislocalization (blue bars) for each experiment. Wild-type and *Sirt1*<sup>-/-</sup> percentages are shown on the left of the graph. The percentage of colocalization/mislocalization of the transfected cells are represented on the right of the graph. These experiments, which involved the quantification of 800 Flag-positive cells for each condition, were reproduced in triplicate and error bars (s.d.) are included. Representative immunofluorescence images of these experiments are shown in panel **B**, **i**–**t**.



distribution was slightly more punctuate in the mutant case (Fig. 4A, compare f and h). However, the distribution of the SIRT1 target, H4K16ac, was altered greatly. Although normally distributed to euchromatin (Fig. 4A, j), H4K16ac was now seen to invade constitutive heterochromatin, being almost exclusively located there in the case of *Suv*<sup>-/-</sup> cells (Fig. 4A, l). None of the other non-heterochromatin-associated modifications tested (H4K12ac, H3K9ac or H3K4me3) was found to invade heterochromatin in these cells (Supplementary Fig. 6a).

We next gauged SUV39H1 activity in the absence of SIRT1. *Sirt1*<sup>-/-</sup> MEFs showed a complete loss of SUV39H1-dependent H3K9me3 in heterochromatic regions (Fig. 4B, D) in more than 50% of the cells studied, in comparison to the wild-type case (Fig. 4C). This is in agreement with previous findings in which SIRT1 loss by RNA interference led to decreased levels of H3K9me3 (ref. 3) and in which Ctr4-mediated methylation of H3K9me3 in constitutive heterochromatin required the presence of Sir2 in *S. pombe*<sup>6</sup>. Moreover, the localization of heterochromatin protein 1 (HP1) to the same loci was also abrogated in *Sirt1*<sup>-/-</sup> MEFs (Fig. 4B–D). However, in contrast to the case of *S. pombe*, the loss of SIRT1 or SUV39H1 in the mammalian system studied here did not result in an invasion of H3K9ac in those regions that exhibited a loss of H3K9me3 (Supplementary Fig. 6b). This indicates that mechanisms engendering heterochromatin formation may not be completely conserved from yeast to humans. Of note, the antibodies against H3K9me3 were derived using a branched methylated peptide, and the detection of H3K9me3 is SUV39H1-dependent (Fig. 4A, compare b with d). This effect is directly dependent on SIRT1, because expression of exogenous SIRT1 rescued H3K9me3 mislocalization in heterochromatic regions (Fig. 4B, i–l, E). This SIRT1-mediated rescue depends on both its activity and the presence of its N-terminal domain (Fig. 4B, m–t, E), supporting a direct functional relationship between SIRT1 and SUV39H1. In addition, the high degree of identity between SUV39H1 and its close relative SUV39H2, together with our results obtained with the *Suv*<sup>-/-</sup> cells, may suggest also a SIRT1 link to SUV39H2, the function of which is not completely understood. Further work should shed a light on this point.

This report demonstrates the direct and functional interrelationship between two major enzymes that implement chromatin structural changes through histone modifications during the formation of heterochromatin. This underscores that the enzymes responsible for pivotal histone modifications do not function in isolation, but are co-regulatory. SIRT1 exhibits a multi-pronged role in promoting conditions favourable to transcriptional repression by deacetylating specific histone proteins, recruiting histone H1b and, as shown here, modulating SUV39H1 activity, resulting in the accumulation of another repressive chromatin mark. This streamlining of the participants that alter chromatin structure is just beginning to be appreciated and probably operates during other regulatory events that evolved to ensure this process.

## METHODS SUMMARY

**Immunofluorescence.** The MEFs used in this work were fixed in 4% paraformaldehyde for 10 min at room temperature (22 °C). Membrane permeabilization was achieved by incubation in buffer A (0.1% sodium azide PBS, 0.5% Triton-X, 0.5–1% BSA) for 10 min at room temperature, as described previously<sup>13</sup>; this was followed by incubation with antibodies against H3K9me3, mouse SIRT1, HP1 $\alpha$  and H4AK12 (Upstate); H3K9ac and H3K4me3 (Abcam); and H4K16ac (Serotec).

**Mapping of acetylated residue in SUV39H1.** Gel-resolved proteins from nicotinamide-treated and untreated (control) samples were digested with trypsin, batch-purified on a reversed-phase micro-tip, and analysed by matrix-assisted laser desorption/ionization reflectron time-of-flight (MALDI-reTOF) mass spectrometry (MS) (UltraFlex TOF/TOF, Bruker) for peptide mass fingerprinting, as described<sup>14</sup>. This served to confirm the identity of the proteins

and to locate differences between the tryptic peptide maps of the modified (acetylated) and unmodified forms. Differential peak *m/z* values were matched to the protein sequence provided, allowing for the probable presence of acetyl groups (+42 Da). Mass-spectrometric sequencing of the putative acetylated peptides (with or without methionine oxidation) was then carried out by MALDI-TOF/TOF MS/MS analysis using the UltraFlex instrument in 'LIFT' mode. Fragment ion spectra, derived from averaging 2,000 laser shots, were inspected for a, b and y ions to compare with the computer-generated fragment ion series of the predicted tryptic peptides to locate the exact or approximate position of the acetylated amino acid.

**HKMT assays and immunoprecipitations.** The assays were performed as described previously<sup>15</sup>. Immunoprecipitations were performed using Flag agarose (Sigma) and agarose-crosslinked Myc antibody (Cell Signaling).

**Cells and treatments.** MEFs were generated as described<sup>2,16,17</sup>. Human 293 cells were treated with nicotinamide and/or TSA under conditions described previously<sup>3</sup>. Transfections were performed using Lipofectamin 2000 (Invitrogen) according to the manufacturer's instructions.

Received 26 July; accepted 17 September 2007.

- Craig, J. M. Heterochromatin—many flavours, common themes. *Bioessays* **27**, 17–28 (2005).
- Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593–599 (2000).
- Vaquero, A. *et al.* Human SirT1 interacts with histone H1 and promotes formation of facultative heterochromatin. *Mol. Cell* **16**, 93–105 (2004).
- Imai, S., Armstrong, C. M., Kaeberlein, M. & Guarente, L. Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. *Nature* **403**, 795–800 (2000).
- Tachibana, M. *et al.* G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev.* **16**, 1779–1791 (2002).
- Shankaranarayana, G. D., Motamedi, M. R., Moazed, D. & Grewal, S. I. Sir2 regulates histone H3 lysine 9 methylation and heterochromatin assembly in fission yeast. *Curr. Biol.* **13**, 1240–1246 (2003).
- Sewalt, R. G. *et al.* Selective interactions between vertebrate polycomb homologs and the SUV39H1 histone lysine methyltransferase suggest that histone H3–K9 methylation contributes to chromosomal targeting of Polycomb group proteins. *Mol. Cell Biol.* **22**, 5539–5553 (2002).
- Sauve, A. A., Wolberger, C., Schramm, V. L. & Boeke, J. D. The biochemistry of sirtuins. *Annu. Rev. Biochem.* **75**, 435–465 (2006).
- Monneret, C. Histone deacetylase inhibitors. *Eur. J. Med. Chem.* **40**, 1–13 (2005).
- Bitterman, K. J., Anderson, R. M., Cohen, H. Y., Latorre-Esteves, M. & Sinclair, D. A. Inhibition of silencing and accelerated aging by nicotinamide, a putative negative regulator of yeast Sir2 and human SIRT1. *J. Biol. Chem.* **277**, 45099–45107 (2002).
- Min, J., Zhang, X., Cheng, X., Grewal, S. I. & Xu, R. M. Structure of the SET domain histone lysine methyltransferase Ctr4. *Nature Struct. Biol.* **9**, 828–832 (2002).
- Melcher, M. *et al.* Structure–function analysis of SUV39H1 reveals a dominant role in heterochromatin organization, chromosome segregation, and mitotic progression. *Mol. Cell Biol.* **20**, 3728–3741 (2000).
- Vaquero, A. *et al.* SirT2 is a histone deacetylase with preference for histone H4 Lys16 during mitosis. *Genes Dev.* **20**, 1256–1261 (2006).
- Winkler, G. S. *et al.* Isolation and mass spectrometry of transcription factor complexes. *Methods* **26**, 260–269 (2002).
- Nishioka, K. *et al.* PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin. *Mol. Cell* **9**, 1201–1213 (2002).
- Peters, A. H. *et al.* Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107**, 323–337 (2001).
- Cheng, H. L. *et al.* Developmental defects and p53 hyperacetylation in Sir2 homolog (SIRT1)-deficient mice. *Proc. Natl Acad. Sci. USA* **100**, 10794–10799 (2003).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank T. Jenuwein for sharing expression plasmids for wild-type and mutant versions of SUV39H1 and SUV39H1 MEF cells (wild type and -/-); F. W. Alt for SIRT1 MEF cells (wild type and -/-); P. Trojer and other members of the Reinberg laboratory for discussions; L. Vales for comments on the manuscript; K. Cabane for technical assistance; and L. Lacomis for help with mass spectrometric analysis. This work was supported by the NIH and HHMI (D.R.) and the NCI Cancer Center (P.T.).

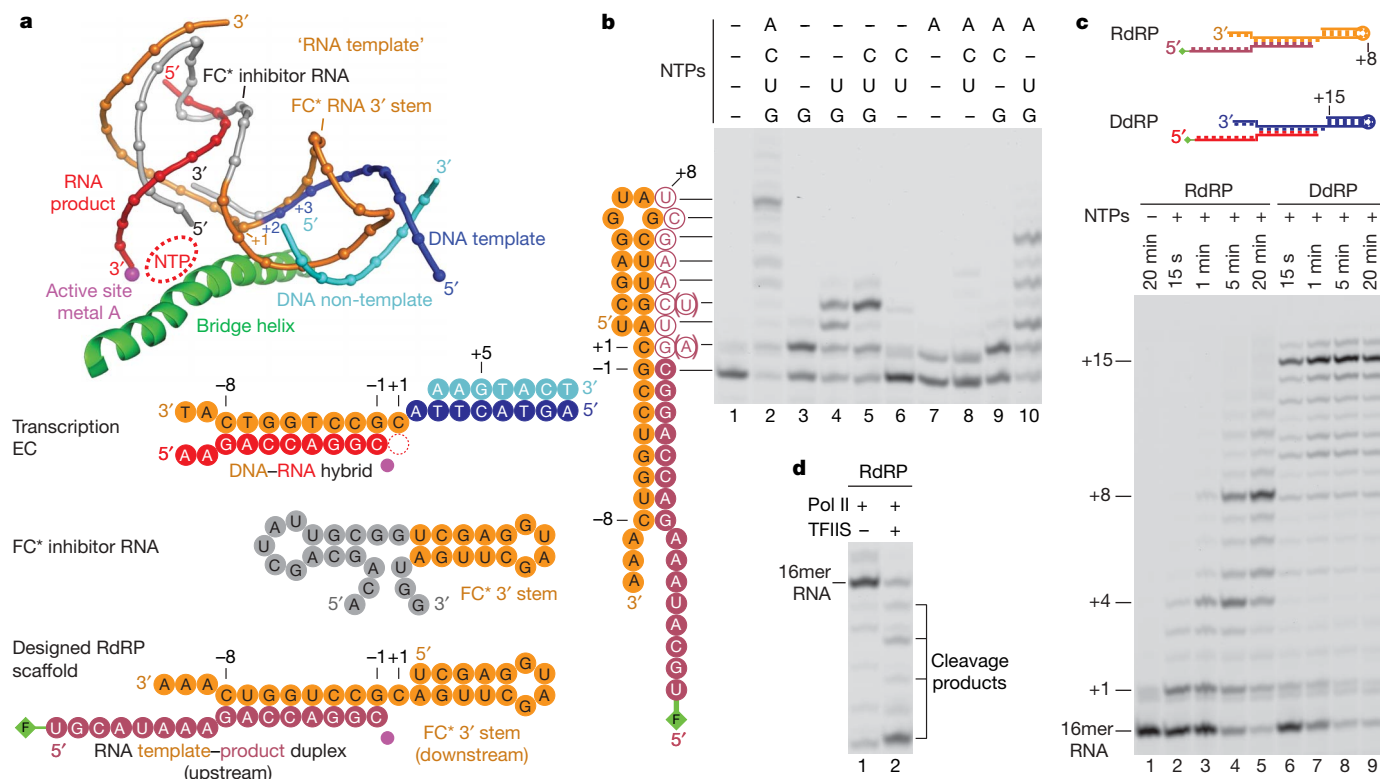
**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to D.R. ([reinbd01@med.nyu.edu](mailto:reinbd01@med.nyu.edu)).

# Molecular basis of RNA-dependent RNA polymerase II activity

Elisabeth Lehmann<sup>1</sup>, Florian Brueckner<sup>1</sup> & Patrick Cramer<sup>1</sup>

RNA polymerase (Pol) II catalyses DNA-dependent RNA synthesis during gene transcription. There is, however, evidence that Pol II also possesses RNA-dependent RNA polymerase (RdRP) activity. Pol II can use a homopolymeric RNA template<sup>1</sup>, can extend RNA by several nucleotides in the absence of DNA<sup>2</sup>, and has been implicated in the replication of the RNA genomes of hepatitis delta virus (HDV)<sup>3,4</sup> and plant viroids<sup>5</sup>. Here we show the intrinsic RdRP activity of Pol II with only pure polymerase, an RNA template–product scaffold and nucleoside triphosphates (NTPs). Crystallography reveals the template–product duplex in the site occupied by the DNA–RNA hybrid during transcription.

RdRP activity resides at the active site used during transcription, but it is slower and less processive than DNA-dependent activity. RdRP activity is also obtained with part of the HDV antigenome. The complex of transcription factor IIS (TFIIS) with Pol II can cleave one HDV strand, create a reactive stem-loop in the hybrid site, and extend the new RNA 3' end. Short RNA stem-loops with a 5' extension suffice for activity, but their growth to a critical length apparently impairs processivity. The RdRP activity of Pol II provides a missing link in molecular evolution, because it suggests that Pol II evolved from an ancient replicase that duplicated RNA genomes.



**Figure 1 | RNA-dependent RNA polymerase II activity.** **a**, Structure-based design of the RdRP scaffold. Structures of the complete Pol II EC<sup>7</sup> and the Pol II–FC\* RNA inhibitor complex<sup>9</sup> were superimposed by fitting aspartate loop residues in the active site. The nucleic acids in the active centre of the two complex structures (EC' and FC\* RNA<sup>9</sup>) are shown. The RdRP scaffold derived by combination of a forked RNA template–product duplex with the 3' stem of FC\* RNA is shown at the bottom. The RNA product strands are in red (transcription EC) or raspberry (RdRP EC), and the RNA template strand is in orange. The FAM fluorescent label is shown as a green diamond. The same colour code is used throughout. The 5' stem of FC\* RNA, which is

not used here, is in grey. The nucleotide addition site is denoted +1, and upstream and downstream positions are depicted with negative and positive numbers, respectively. **b**, RNA synthesis with the RdRP scaffold from **a**. Lane 1 shows the fluorescently labelled reactant RNA. In lanes 2–10 the RdRP scaffold was incubated with pure Pol II and different types of NTP (see Methods). UMP misincorporation at position +3 (lane 4) was confirmed by mass spectrometry (not shown). **c**, Time courses of RNA synthesis with the RdRP scaffold (lanes 2–5) and the corresponding DNA template-containing scaffold DdRP (lanes 6–9). **d**, TFIIS-induced 3'-RNA cleavage in the RdRP EC.

<sup>1</sup>Gene Center Munich and Center for Integrated Protein Science (CIPS<sup>M</sup>), Department of Chemistry and Biochemistry, Ludwig-Maximilians-Universität München, Feodor-Lynen-Strasse 25, 81377 Munich, Germany.

Structural studies of the transcription elongation complex (EC) revealed a B-form DNA duplex entering the downstream polymerase cleft, and a hybrid duplex of DNA template with RNA product above the active site<sup>6–8</sup> (Fig. 1a). During RdRP activity, the hybrid site and the downstream cleft are expected to accommodate A-form RNA duplexes. The downstream cleft can accommodate A-RNA, as observed for the 3' stem of the RNA inhibitor FC\* (ref. 9) (Fig. 1a). The FC\* 3' stem overlaps with two DNA template positions downstream of the NTP-binding site (position +1; Fig. 1a), suggesting that an RNA template could enter the active site in a similar manner to DNA.

To test whether Pol II can indeed use an RNA template, we prepared an RNA scaffold that combined the FC\* 3' stem with a putative RNA template–product duplex (RdRP scaffold; Fig. 1a). To monitor potential RNA elongation, the product strand was labelled at the 5' end with 6-carboxyfluorescein (FAM). Incubation of this scaffold with Pol II and NTPs led to RNA elongation (Fig. 1b, lanes 1 and 2), indicating the formation of an active RdRP EC. Most RNA was elongated by eight nucleotides, but shorter and longer products were also observed. Mass spectroscopy of the products was consistent with RNA-templated synthesis (Supplementary Table 1). Incubation of the RdRP EC with subsets of NTPs led to products that generally arose from templated nucleotide addition (Fig. 1b, lanes 3–10), but the absence of the cognate NTP also permitted misincorporation. At position +1, AMP could be incorporated instead of GMP; at position +3, UMP could replace CMP. Misincorporation did not result from the use of RNA as a template, because a corresponding DNA template (scaffold DdRP; Fig. 1c) produced the same product pattern (Supplementary Fig. 1). These results show the intrinsic RdRP activity of Pol II.

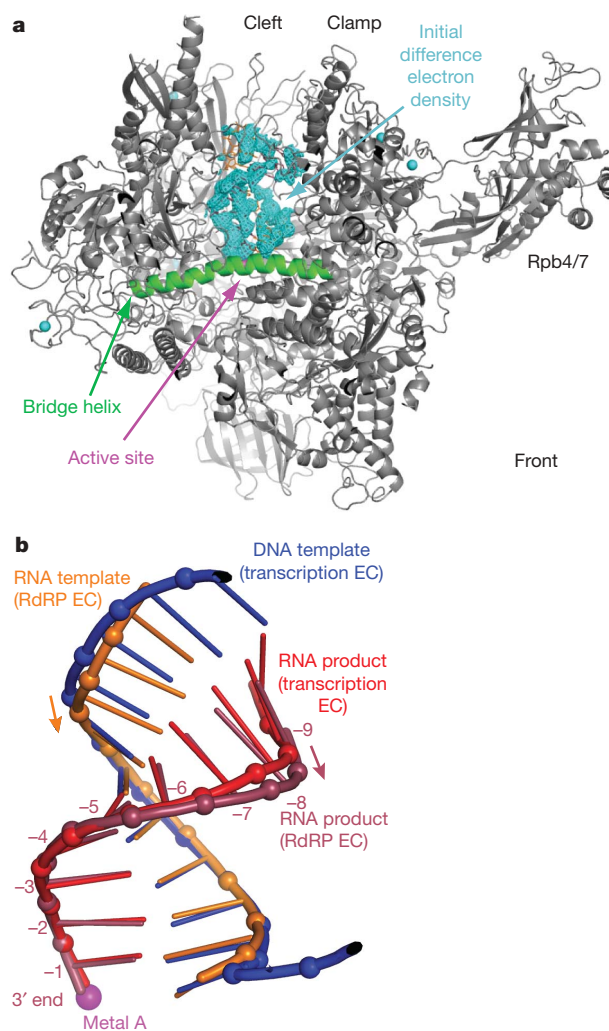
To unravel the structural basis of the RdRP activity, we determined the crystal structure of the complete Pol II bound to an RdRP scaffold. A scaffold with a 5' extension of six nucleotides sufficed to form an active RdRP EC (Supplementary Fig. 2), and enabled crystallographic analysis at 3.8 Å resolution to be made (Fig. 2, Supplementary Fig. 3, Supplementary Table 2, and Methods). The structure revealed the RNA template–product duplex in the site occupied by the DNA–RNA hybrid during transcription. The RNA product 3' end was bound to the catalytic metal ion A (Fig. 2). RNA synthesis occurred at the active site used during transcription, because product RNA was cleaved from the 3' end by TFIIS (Fig. 1d), which stimulates RNA cleavage at the active site in the EC<sup>10</sup>, and in a Pol II–RNA complex<sup>2</sup>. For positions –1 to –5, the structure of the RNA duplex was essentially identical to that of the DNA–RNA hybrid<sup>7</sup>; however, at the upstream positions –6 to –9, changes in phosphate positions of up to 3 Å resulted in a decreased helical rise (Fig. 2b).

To compare Pol II activity on the RdRP scaffold and on a corresponding DNA template scaffold, we conducted a time course experiment (Fig. 1c). Pol II rapidly reached the end of the DNA template but was slower and stopped prematurely on the RNA template (Fig. 1c). This was not due to a different scaffold affinity for Pol II (Supplementary Fig. 4). To investigate the premature stop, we altered the scaffold (Supplementary Fig. 2). When the downstream duplex length was increased or decreased by 1 base pair (bp), the maximum product length changed accordingly (Supplementary Fig. 2b, lanes 3 and 4), suggesting that the uracil at position +9 posed an obstacle to elongation. However, synthesis still stopped at position +8 when this uracil was replaced by cytosine or when the downstream duplex was replaced by a single strand (Supplementary Fig. 2b, d). Thus, Pol II could use single-stranded or double-stranded downstream RNA templates and showed impaired processivity, independently of the nature of downstream RNA.

To investigate the physiological significance of the RdRP activity, we studied a terminal segment of the HDV antigenome (Fig. 3a). In cell extracts, this segment directs RNA synthesis that is sensitive to the Pol II inhibitor  $\alpha$ -amanitin<sup>11,12</sup>. The reaction involved RNA strand

cleavage followed by elongation of the new 3' end<sup>11,12</sup>. The cleaved HDV segment apparently forms an RNA template–product stem-loop and a downstream RNA duplex, which we shortened by replacing a natural bulge with a loop (HDV scaffold; Fig. 3a). Incubation of this scaffold with Pol II and NTPs resulted in RNA synthesis up to the end of the template, although synthesis also stopped prematurely (Fig. 3b, lane 8). RNA synthesis was strictly dependent on the template. Only cognate NMPs were incorporated at positions +1 and +2 (Fig. 3b, lanes 2–6). When UTP was omitted from the reaction, RNA synthesis stopped at position +13 as expected (Fig. 3b, lane 7). RNA synthesis was slow, but some final product appeared within 1 min (Fig. 3d). These *in vitro* data match the evidence for Pol II-dependent HDV replication *in vivo*<sup>3,4</sup>, and argue for the physiological significance of the RdRP activity of Pol II.

The higher processivity of Pol II on the HDV scaffold than on the RdRP scaffold may originate from the different upstream template–product duplex, which forms a stem-loop instead of a forked end (Fig. 3c). We therefore analysed two chimaeric scaffolds that recombine the upstream and downstream regions of the two scaffolds



**Figure 2 | Crystal structure of a Pol II–RdRP complex.** **a**, Ribbon model of Pol II (grey) with an initial, unbiased difference Fourier electron-density map (cyan, contoured at  $2.2\sigma$ ). The map was calculated from protein model phases. It reveals the RNA template–product duplex of scaffold RdRP–ss6 (Supplementary Fig. 2) in the active-centre cleft of Pol II. The bridge helix is in green. The catalytic metal ion A is depicted as a magenta sphere, and  $\text{Zn}^{2+}$  ions as cyan spheres. The view is related to that in Fig. 1a by a  $90^\circ$  rotation around a vertical axis. **b**, Comparison of the RNA template–product duplex in the RdRP EC with the DNA–RNA hybrid duplex in the transcription EC<sup>7</sup>. Protein structures were superimposed by fitting the active-site aspartate loops.



(Fig. 3c). Combination of the RdRP upstream region with the HDV downstream region strongly decreased processivity compared with the HDV scaffold (Fig. 3e). In contrast, combination of the HDV upstream stem-loop with the RdRP downstream region enabled run-off synthesis that was not possible with scaffold RdRP (Figs 1c and 3f). Thus, the HDV stem-loop promoted RdRP processivity. Consistently, the stem-loop alone, containing only a two-nucleotide 5' extension, formed a functional RdRP EC (Fig. 4a).

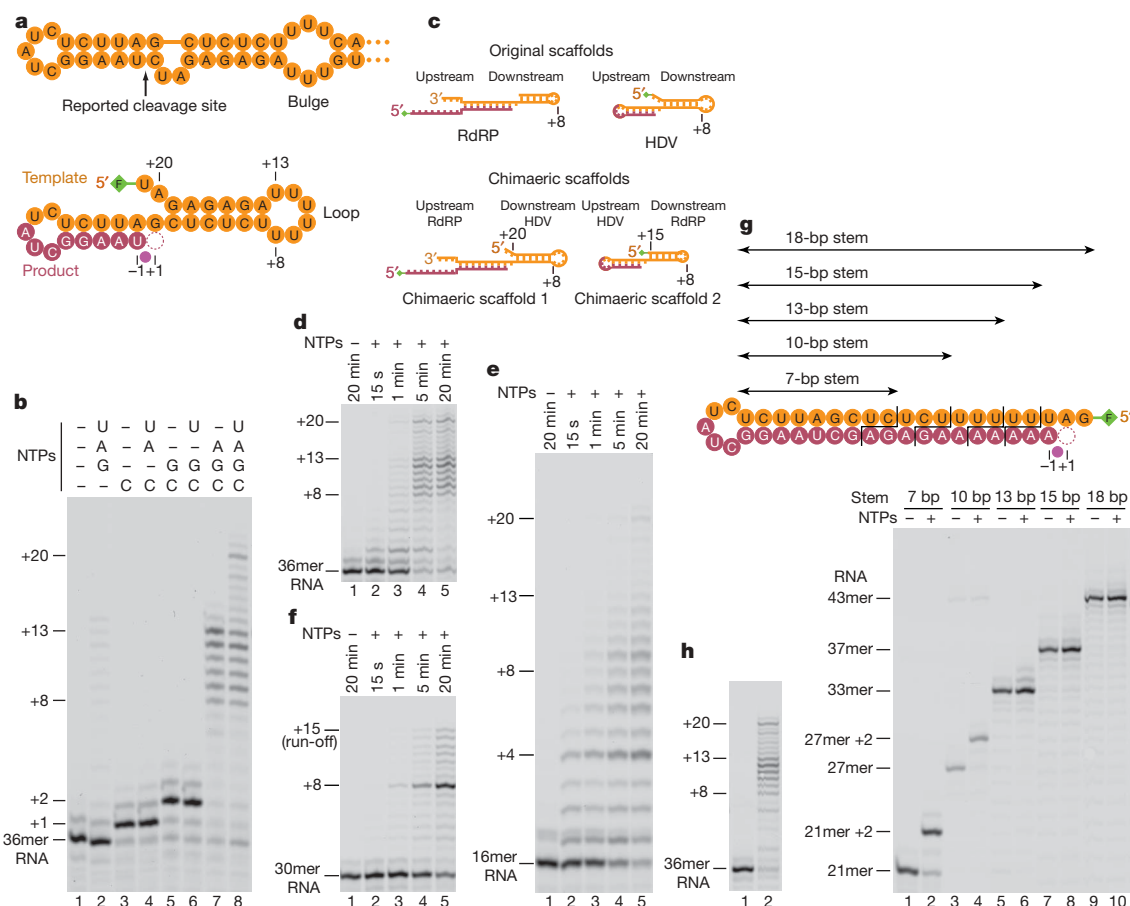
To test whether the RdRP-promoting stem-loop can be formed from the HDV terminal segment *in vitro*, we incubated the segment with pure Pol II–TFIIS complex. This resulted in RNA cleavage at the internal bulge (Fig. 4b). The bulge apparently connected the two RNA duplexes in a flexible way, to enable positioning of the scissile RNA strand at the active site (Fig. 4e). Cleavage produced a 6-bp stem-loop, comparing favourably with the 5-bp stem-loop that forms in extracts<sup>11</sup>. On the addition of NTP, the new 3' end was elongated up to the end of the template (Fig. 4b, lane 7). Consistently, a corresponding 6-bp stem-loop with a 5' extension supported RdRP activity (Fig. 4a). Thus, the Pol II–TFIIS complex can cleave the HDV terminal segment and elongate the new 3' end *in vitro*, which is consistent with models for HDV replication *in vivo*.

The 6-bp stem-loop that forms by HDV RNA cleavage *in vitro* (Fig. 4a) could be structurally revealed in a complex with Pol II (Fig. 4c, d, and Supplementary Table 2). The stem-loop bound the hybrid site, and phosphates at positions –1 to –5 occupied the same locations as in the artificial RdRP EC. However, RNA positions –6 and –7 had an increased helical rise (Fig. 4d), showing that

the hybrid site accommodates various RNA duplexes with upstream ends of different structure. The RNA loop was mobile except for one cytidine residue (Fig. 4a, c). Consistently, a double mutation in the loop did not impair activity (Supplementary Fig. 5). Because base-specific Pol II–RNA contacts are absent in both RdRP EC structures, Pol II recognizes the A-RNA stem rather than a particular RNA sequence.

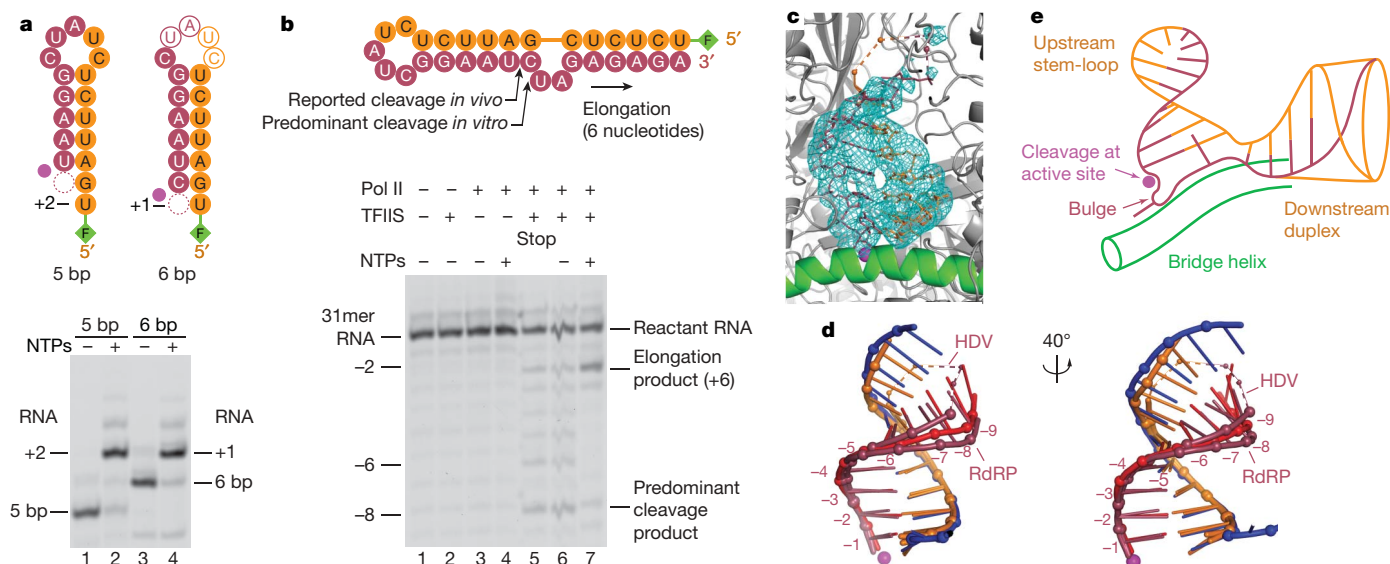
During transcription, upstream template and product strands are separated, whereas the HDV stem-loop probably persists during elongation. Consistently, Pol II readily used stem-loops with 5, 6, 7 or 10 bp as substrates, but not stem-loops with 13, 15 or 18 bp (Figs 3g and 4a). Modelling revealed a clash of the polymerase lid with the RNA stem-loop when it reached a length of 10 or 11 bp. The clash may destabilize the RdRP EC and facilitate stalling. This model explains why elongation with the HDV scaffold partly stalled when the stem-loop reached 13–18 bp (Fig. 3a, b), and why replacement of the uracil at the downstream position +8 did not influence stalling (Fig. 3h). The model also explains stalling of the artificial RdRP ECs when a persistent template–product duplex would reach a length of 16 bp (Fig. 1b and Supplementary Fig. 2). Similarly, a persistent DNA–RNA hybrid stalls the transcription EC<sup>13–15</sup>. The limited RdRP processivity *in vitro* is apparently overcome during HDV replication *in vivo* by binding of the HDV-encoded elongation-stimulatory delta antigen to the polymerase clamp<sup>11,12,16</sup>. Clamp movements and/or RNA repositioning would enable the exit of a persistent stem-loop from the cleft.

Our results also explain the previously observed apparently non-templated RNA elongation in a Pol II–RNA complex<sup>2</sup>. The RNA used



**Figure 3 | RNA synthesis with HDV RNA-derived scaffolds** **a**, Schematic representation of the terminal segment of the HDV antigenome (top) and the HDV-derived scaffold (bottom). **b**, Pol II-dependent RNA synthesis with HDV scaffold. Lane 1 shows the fluorescently labelled reactant RNA. Lanes 2–8 show the elongation products after incubation with various subsets of NTPs as indicated. **c**, Design of two chimaeric scaffolds that recombine the

upstream and downstream regions of the artificial RdRP scaffold and the HDV-derived scaffold. **d–f**, Time courses of RNA synthesis with HDV scaffold (**d**) and with the chimaeric scaffolds 1 (**e**) and 2 (**f**). **g**, HDV-derived stem-loops of different lengths serve as RdRP scaffolds. **h**, RNA synthesis with an HDV scaffold with U in position +8 mutated to C.



**Figure 4 | Mechanism of HDV replication initiation.** **a**, HDV-derived terminal stem-loops consisting of 5 or 6 bp enable templated incorporation of the next nucleotide(s). **b**, Pure Pol II-TFIIS complex cleaves the HDV antigenome terminal segment and elongates the newly formed 3' end on the addition of NTP. For cleavage, Pol II-scaffold complex (625 nM) was incubated with TFIIS (1.25 μM) in transcription buffer for 60 min at 28 °C. For elongation of the cleavage product, the reaction mixture was incubated with 1 mM NTPs at 28 °C for 20 min. For lane 5, the cleavage reaction was stopped after 60 min ('Stop'). **c**, Difference electron density omit map for the 6-bp HDV stem-loop bound to the hybrid site of Pol II (calculated with

for these studies bound Pol II and was cleaved on the addition of TFIIS; cleavage products were elongated by a few nucleotides<sup>2</sup>. We predict that this RNA formed an 11-bp stem-loop in the hybrid site, and a bulge or a 3' tail at the active site (Supplementary Fig. 6). TFIIS-stimulated cleavage then created a new 3' end at the active site, which was elongated in a templated manner, until a critical template-product length was reached. Consistently, an 11-bp stem-loop resembling a cleaved species<sup>2</sup> supported templated nucleotide incorporation (Supplementary Fig. 6). Our results also explain previously observed cleavage and limited re-extension of a short RNA stem-loop by bacterial RNA polymerase<sup>17</sup>. Thus, earlier observations<sup>2,17</sup> are apparently examples of the RdRP mechanism described here.

The RdRP activity of Pol II provides a missing link in molecular evolution, because it supports the prediction that an ancestor of Pol II was the enzyme that replicated early RNA genomes<sup>18</sup>. This is supported by the recent findings that the related bacterial RNA polymerase supports not only RNA synthesis with the non-coding 6S RNA as template<sup>19,20</sup> but also RNA-facilitated RNA cleavage<sup>21</sup>. The early RNA replicase apparently evolved to accept DNA as a template during the transition from RNA to DNA genomes. The early replicase was apparently a common ancestor also of distinct single-subunit RdRP enzymes, because these show a Pol II-like core protein fold and active site<sup>22</sup>. It is possible that the ancient RdRP activity of Pol II, which still replicates HDV, is still used in certain cellular processes, because many organisms lack dedicated single-subunit RdRPs. Given the slow nature and the impaired processivity of the RdRP activity *in vitro*, such processes would require stimulating factors or would be restricted to the generation of smaller RNAs.

## METHODS

Endogenous *Saccharomyces cerevisiae* core Pol II was purified as described<sup>23</sup> except that the anion-exchange step was omitted. Recombinant Rpb4/7 and TFIIS were prepared as described<sup>10,24</sup>. Pol II-scaffold complexes were assembled by incubating pure Pol II with two molar equivalents of annealed nucleic acids in transcription buffer as described<sup>24</sup>, at 20 °C for 30 min. Unless stated otherwise,

protein phases only, contoured at 3.0σ). The disordered loop is indicated with a dashed line. The view is as in Fig. 2a. **d**, Superposition of the RNA template-product duplex in the HDV EC and the RdRP EC (Fig. 2) on the DNA-RNA hybrid duplex in the transcription EC<sup>7</sup>. Protein structures were superimposed by fitting the active-site aspartate loops. **e**, Model of initial interaction of the HDV antigenome terminal segment with the Pol II-TFIIS complex. The stem-loop is placed in accordance with the crystal structure (c, d) and the downstream duplex in accordance with the location of the FC\* RNA 3' stem. We predict that the HDV bulge passes the bridge helix and active site, where cleavage occurs.

Pol II-scaffold complex (330 nM) was incubated with NTPs (1 mM) at 28 °C for 20 min for RNA extension assays, or with TFIIS (660 nM) for 60 min for RNA cleavage assays. Reactions were stopped by incubation with an equal volume of 8 M urea, 2 × TBE (Tris-borate-EDTA) for 5 min at 95 °C. The FAM-labelled RNA products were separated by denaturing gel electrophoresis (1 pmol of RNA per lane, 0.4-mm 15–20% polyacrylamide gels containing 7 M urea, at 50–55 °C) and revealed with a Typhoon 9400 scanner (GE Healthcare). The crystal structure of the Pol II complex with scaffold RdRP-ss6 (Supplementary Fig. 2) was determined essentially as described<sup>7,24</sup>. Core Pol II was incubated with two molar equivalents of the scaffold and five molar equivalents of Rpb4/7 in 5 mM HEPES pH 7.25, 40 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 10 μM ZnCl<sub>2</sub>, 10 mM dithiothreitol. The complex was purified by size-exclusion chromatography, concentrated to 4 mg ml<sup>-1</sup>, and crystallized at 22 °C in hanging drops containing 2 μl of sample and 1 μl of reservoir solution (200 mM ammonium acetate, 150 mM magnesium acetate, 50 mM HEPES pH 7.0, 5% poly(ethylene glycol) 6000, 5 mM TCEP (tris (2-carboxyethyl) phosphine hydrochloride)). The HDV stem-loop was co-crystallized with Pol II under the same conditions. RNA scaffold (10 μM) was added to the cryo solutions. For crystallographic details see Supplementary Table 2. Diagrams were prepared with PyMOL (DeLano Scientific).

Received 18 July; accepted 21 September 2007.

- Dezelee, S., Sentenac, A. & Fromageot, P. Role of deoxyribonucleic acid-ribonucleic acid hybrids in eukaryotes. *J. Biol. Chem.* **249**, 5978–5983 (1974).
- Johnson, T. L. & Chamberlin, M. J. Complexes of yeast RNA polymerase II and RNA are substrates for TFIIS-induced RNA cleavage. *Cell* **77**, 217–224 (1994).
- Lai, M. M. C. RNA replication without RNA-dependent RNA polymerase: surprises from hepatitis delta virus. *J. Virol.* **79**, 7951–7958 (2005).
- Taylor, J. M. Replication of human hepatitis delta virus: recent developments. *Trends Microbiol.* **11**, 185–190 (2003).
- Rackwitz, H. R., Rohde, W. & Sanger, H. L. DNA-dependent RNA polymerase II of plant origin transcribes viroid RNA into full-length copies. *Nature* **291**, 297–301 (1981).
- Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292**, 1876–1882 (2001).
- Kettenberger, H., Armache, K.-J. & Cramer, P. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol. Cell* **16**, 955–965 (2004).

8. Westover, K. D., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell* **119**, 481–489 (2004).
9. Kettenberger, H. *et al.* Structure of an RNA polymerase II–RNA inhibitor complex elucidates transcription regulation by noncoding RNAs. *Nature Struct. Mol. Biol.* **13**, 44–48 (2006).
10. Kettenberger, H., Armache, K.-J. & Cramer, P. Architecture of the RNA polymerase II–TFIIS complex and implications for mRNA cleavage. *Cell* **114**, 347–357 (2003).
11. Filipovska, J. & Konarska, M. M. Specific HDV RNA-templated transcription by pol II *in vitro*. *RNA* **6**, 41–54 (2000).
12. Yamaguchi, Y. *et al.* Stimulation of RNA polymerase II elongation by hepatitis delta antigen. *Science* **293**, 124–127 (2001).
13. Kireeva, M. L., Komissarova, N. & Kashlev, M. Overextended RNA:DNA hybrid as a negative regulator of RNA polymerase II processivity. *J. Mol. Biol.* **299**, 325–335 (2000).
14. Touloukhonov, I. & Landick, R. The role of the lid element in transcription by *E. coli* RNA polymerase. *J. Mol. Biol.* **361**, 644–658 (2006).
15. Naryshkina, T., Kuznedelov, K. & Severinov, K. The role of the largest RNA polymerase subunit lid element in preventing the formation of extended RNA–DNA hybrid. *J. Mol. Biol.* **361**, 634–643 (2006).
16. Yamaguchi, Y., Mura, T., Chanarat, S., Okamoto, S. & Handa, H. Hepatitis delta antigen binds to the clamp of RNA polymerase II and affects transcriptional fidelity. *Genes Cells* **12**, 863–875 (2007).
17. Kashlev, M. & Komissarova, N. Transcription termination: primary intermediates and secondary adducts. *J. Biol. Chem.* **277**, 14501–14508 (2002).
18. Poole, A. M. & Logan, D. T. Modern mRNA proofreading and repair: clues that the last universal common ancestor possessed and RNA genome? *Mol. Biol. Evol.* **22**, 1444–1455 (2005).
19. Wassarman, K. M. & Saecker, R. M. Synthesis-mediated release of a small RNA inhibitor of RNA polymerase. *Science* **314**, 1601–1603 (2006).
20. Gildehaus, N., Neusser, T., Wurm, R. & Wagner, R. Studies on the function of the riboregulator 6S RNA from *E. coli*: RNA polymerase binding, inhibition of *in vitro* transcription and synthesis of RNA-directed *de novo* transcripts. *Nucleic Acids Res.* **35**, 1885–1896 (2007).
21. Zenkin, N., Yuzenkova, Y. & Severinov, K. Transcript-assisted transcriptional proofreading. *Science* **313**, 518–520 (2006).
22. Salgado, P. S. *et al.* The structure of an RNAi polymerase links RNA silencing and transcription. *PLoS Biol.* **4**, e434 (2006).
23. Armache, K.-J., Kettenberger, H. & Cramer, P. Architecture of the initiation-competent 12-subunit RNA polymerase II. *Proc. Natl Acad. Sci. USA* **100**, 6964–6968 (2003).
24. Brueckner, F., Hennecke, U., Carell, T. & Cramer, P. CPD damage recognition by transcribing RNA polymerase II. *Science* **315**, 859–862 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank U. Hennecke and members of the Cramer laboratory for help, and J. Doudna, K. Förstemann, G. Meister and R. Schroeder for discussions. This work was supported by the Deutsche Forschungsgemeinschaft, the SFB646, the Nanoinitiative Munich, the Elitenetzwerk Bayern, the EU grant 3D repertoire, and the Fonds der Chemischen Industrie.

**Author Information** The coordinates and structure factors for the RdRP EC and the HDV EC have been deposited in the Protein Data Bank under accession codes 2R92 and 2R93, respectively. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to P.C. ([cramer@LMB.uni-muenchen.de](mailto:cramer@LMB.uni-muenchen.de)).



## CORRIGENDUM

doi:10.1038/nature06439

**Positive darwinian selection at the imprinted  
*MEDEA* locus in plants**

Charles Spillane, Karl J. Schmid, Sylvia Laouëillé-Duprat,  
Stéphane Pien, Juan-Miguel Escobar-Restrepo, Célia Baroux,  
Valeria Gagliardini, Damian R. Page, Kenneth H. Wolfe  
& Ueli Grossniklaus

*Nature* 448, 349–352 (2007)

---

We inadvertently omitted to submit some of the sequences we used to GenBank. The complete set of GenBank numbers is: DQ975464, DQ975465, EU073602 to EU073616, EU073618 to EU073661, and EU139161 to EU139163.

# A better mousetrap

The appliance of science.

**Mike Resnick**

Men have always talked about building a better mousetrap, just the way they talk about a car that runs on water rather than gasoline, or nuclear fission that doesn't have any harmful by-products. But it wasn't until they reopened the Heisenberg Space Station out between Europa and Callisto that they realized they *really* needed a better mousetrap.

The first team of scientists — four men and two women — docked their ship there on 2 November 3014 AD, at exactly 7:43 p.m. H.T. (Heisenberg time). They buttressed the hatch up against the entrance to the station, sealed it, then opened both doors and stepped into the station, the first humans to do so in more than 900 years.

Exactly 43 seconds later, one of the women screamed, and the other jumped onto a chair that was bolted into the floor. Three of the men started cursing, and the fourth, a wimpy little fellow, fainted dead away.

It seemed that some of the station's inhabitants were waiting for them. They'd been there nine long centuries, and were glad to have some company. Having just eaten the last of the huge stores of preserved food that prior crews had laid in, they were even happier to have a new source of protein.

"What are they?" asked the wimpy scientist when they woke him up.

"Mice," said the nuclear physicist. "Or maybe rats."

"I don't care what they are!" said the roboticist from atop her chair. "Get them away from me!"

"No problem," said the biochemist. "I'll whip up a fast-acting poison and lay it out for them."

At which point the wimpy scientist fainted again.

So the biochemist mixed up the poison, and left it out for the mice, and the crew went about setting up their workstations, ate dinner and went to bed, expecting to find a few hundred dead mice in the morning.

What they found were some plump mice, happily licking their chops and looking for more poisoned bait.

"They've evolved," said the biochemist. "They've obviously developed an immunity to poison. I suppose we'll just have to find some other way to kill them."

"I know just the thing," said the nanotechnologist. "I'll design a mechanical

microbe that will invade their systems and attack them from the inside, and I'll slip it in some cheese."

The mice came, and they saw, and they ate — and they came back the next morning looking for more.

"I don't understand it," said the nanotechnologist. "Those microbes would kill any one of us. Why didn't they kill the mice?"

No one knew, so they captured one of the mice, drew blood samples, stomach samples, gene samples and still had no answer. The best suggestion came from the biochemist, who theorized that their forced evolution had created an internal environment so hospitable to microbes, even engineered ones, that the microbes ignored their programming and set up shop in the mice's intestines.

The roboticist tried next. She created an army of tiny metal warriors and sent them forth to do battle under chairs and beds, inside bulkheads, wherever the mice were hiding.

That was when they learned that the mice had evolved mentally as well as physically, and that their commanders were far superior at warfare to the roboticist, who had programmed her metal army. The robots were outflanked and outmanoeuvred, and finally surrendered only 17 hours into the battle.

The nuclear physicist didn't do much better with his jerrybuilt disintegrator ray. The mice were impervious to it, and the only harm it did was to two bathrooms and the coffee-maker in the galley.

"Well, I'm all out of ideas," said the biochemist.

"The dirty little swine have beaten us at every turn," muttered the nuclear scientist.

"Idiots!" said the wimpy little scientist disgustedly.

"The mice?"

"No," he said. "I was referring to my colleagues."

"You should talk!" snapped the roboticist. "All you ever do is faint."

"I have never denied my limitations," said the scientist, "though it is thoughtless

of you to refer to them. Just for that, I've a good mind not to solve your problem."

"So you think you're the one who can build a better mousetrap?" she said sardonically.

"Most certainly."

"Even though they've withstood poison, microbes, military robots and disintegrator rays?"

"Yes."

"Okay, hot-shot. What will you need?"

"Just a little help from our geneticist."

"And nothing else?"

"Not at the moment," said the scientist.

So they left him and the geneticist alone for a month and tried not to notice all the damage the mice were doing. And then one day the scientist announced that the better mousetrap had been created and was ready to perform its function.

The others all snickered at him.

"That's *it*?" asked the nanotechnologist when he displayed it. "That's the better mouse-

trap that we've been waiting for all month?"

"You don't really think something this primitive is going to work, do you?" demanded the biochemist.

"Oh, ye of little faith," muttered the wimpy scientist.

They all laughed. (Well, they laughed at Newton and Einstein too.)

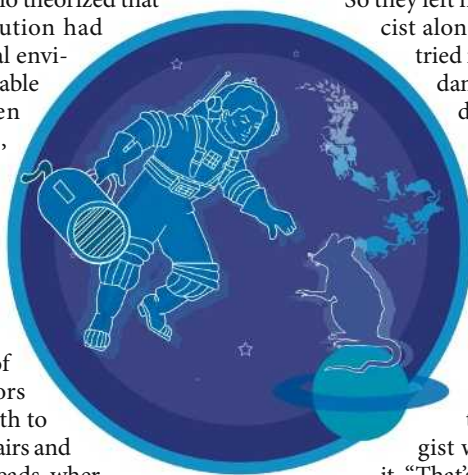
Within a week every mouse on the station had been eliminated, including three that had somehow migrated onto the docked ship. It had been swift, efficient and devastating.

"Who'd have believed it?" said the roboticist as they all gathered around the better mousetrap.

"Where did you ever hear about something like this?" asked the nanotechnologist.

"Sometimes you have to read books that aren't exclusive to your field of study," answered the scientist.

"Meow," agreed the better mousetrap. ■  
**Mike Resnick is the all-time leading award winner, living or dead, for short science fiction, according to *Locus*, the trade journal of the science-fiction field.**



JACEY